Aalto University
School of Science
Master's Programme in Information Networks

Kasper Kylmälä

Recognising Barriers for Refugee Employment With Logistic Regression:

Analysis of refugee background information

Master's Thesis
Espoo, October 16, 2019

Supervisor: Risto Sarvas, PoP
Advisor: Elisa Vepsäläinen, MSSc

Aalto University
School of Science

Aalto University
School of Science
Master's Programme in Information Networks

ABSTRACT OF
MASTER'S THESIS

Author:	Kasper Kylmälä		
Title:	Recognising Barriers for Refugee Employment With Logistic Regression: Analysis of refugee background information		
Date:	October 16, 2019	**Pages:**	vi + 93
Major:	Information Networks	**Code:**	SCI3047
Supervisor:	Risto Sarvas, PoP		
Advisor:	Elisa Vepsäläinen, MSSc		

Integration of refugees and asylum seekers has become a relevant topic all over Europe after the beginning of the refugee crisis in 2015. How to make the integration process more efficient for the government and better for the refugees is a general issue in politics, academic research, and between third sector organisations. Rapid employment of refugees has risen as an alternative model for solving this issue.

Focusing on refugee employment seems to make both the refugees and the host nations better off. There are less unemployed people to take care of and the nations experience an increase in labour force. The refugees benefit by having a place to belong in a new society and earning their own living. Having refugees working instead of depending on welfare inhibits the view that refugees are a burden, but makes it easier to view them as an active, contributing part of society.

This study aims to figure out through quantitative means which background factors have a significant effect on refugee employment. This is done through statistical outlooks into the refugee population and a series of logistic regression models. The study concludes to a model which examines the importance of different factors and their relationship to refugees' employment chances. The model's validity is tested through triangulation with qualitative surveys.

Work experience and English language are concluded as being the most important factors determining one's chance for employment, but one's Finnish language, education level, and profession are important as well. Refugee women's worse position in the labour market is mostly explained by women having less work experience than men.

This study proposes changes in Finland's refugee policies based on statistical evidence.

Keywords:	Data science, data mining, analytics, logistic regression, refugees, rapid employment, asylum seekers
Language:	English

Aalto-yliopisto
Perustieteiden korkeakoulu
Informaatioverkostojen maisteriohjelma

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Kasper Kylmälä

Työn nimi:
Pakolaisten työllistymisen esteiden tunnistaminen logistisella regressiolla: Pakolaisten taustatietojen analysointi

Päiväys:	16. lokakuuta 2019	**Sivumäärä:**	vi + 93
Pääaine:	Informaatioverkostot	**Koodi:**	SCI3047
Valvoja:	Työelämäprofessori Risto Sarvas		
Ohjaaja:	Valtiotieteiden maisteri Elisa Vepsäläinen		

Pakolaisten ja turvapaikanhakijoiden integroitumisesta on tullut merkittävä puheenaihe kaikkialle Euroopassa pakolaiskriisin alettua vuonna 2015. Integraatioprosessista, kuinka sen saisi toteutumaan helpommin ja paremmin, on tullut yleinen aihe sekä politiikassa, akateemisessa keskustelussa että pakolaisjärjestöjen kesken. Pakolaisten pikaisesta työllistämisestä tuli nopeasti mielenkiintoinen uusi malli ongelman ratkaisemiseksi.

Keskittyminen pakolaisten työllistämiseen näyttää hyödyntävän molempia osapuolia: sekä pakolaisia että vastaanottajavaltioita. Vastaanottajavaltioiden ei tarvitse tukea työttömiä pakolaisia ja saavat samalla talouteensa lisää työvoimaa. Pakolaiset hyötyvät löytämällä merkityksellisen yhteisön uudessa kotimaassaan ja saavat tienattua elantonsa itse. Sen sijaan, että yhteiskunta tukisi työttömiä pakolaisia, vaan aktiivisesti työllistäisi heitä, tekee hankalammaksi nähdä pakolaisia ainoastaan yhteiskunnan taakkana. Se mahdollistaa heidän roolinsa aktiivisena osana yhteiskuntaa.

Tämä tutkimus pyrkii selvittämään kvantitatiivisin keinoin, mitkä taustatekijät vaikuttavat merkittävästi pakolaisten työllistymiseen. Tämä toteutetaan tilastollisilla katsauksilla pakolaispopulaatioon ja logistisilla regressiomalleilla. Tämä tutkimus tuottaa mallin jolla on mahdollista tarkkailla eri tekijöiden tärkeyttä ja suhdetta pakolaisten työllistymistodennäköisyyksiin. Mallin pätevyyttä testataan trianguloimalla tuloksia kvalitatiivisilla kyselyillä.

Työkokemus ja englannin kielen taito todetaan tärkeimmiksi tekijöiksi määrittämään työllistymisen todennäköisyyttä, mutta suomen kielen taito, koulutustaso ja ammatti ovat myös tärkeitä. Pakolaisnaisten heikompi asema työmarkkinoilla selittyy suurimmilta osin naisten vähäisemmällä työkokemuksella.

Tämä tutkimus ehdottaa muutoksia Suomen pakolaispolitiikassa, perustuen tilastollisiin todisteisiin.

Asiasanat:	Datatiede, datan louhinta, analytiikka, logistinen regressio, pakolaiset, pikainen työllistyminen, turvapaikanhakijat
Kieli:	Englanti

Acknowledgements

This thesis project was arguably one of the most intense individual projects I have ever undertaken. It was a journey of learning completely new things on my own, figuring out new knowledge, explaining complex phenomena to other people, and of spotting my own strengths more clearly, as well as of tiredness, frustration, and learning through failure. A valuable experience in life with both a darker side and a luminous silver lining.

I would first of all like to thank Startup Refugees and my instructor Elisa Vepsäläinen for offering me the awesome opportunity of being the analyst to research the vast collection of refugee data. Someone had to do it and it happened to be me, which I am grateful for, since the experience will certainly affect my life by opening various doors in the future. Also big thanks to my supervisor Risto Sarvas for the guidance that helped me perfect this thesis to its final form, but especially for pointing me in the right direction that eventually led me to Startup Refugees and their research project.

Last but definitely not the least, I want to thank my friends: Markus and Henri for their overall guidance in the thesis process, Jesse for remembering me to have specific focus in the research, as well as Ville and Väinö for inspiring conversations about mathematical methods. I also want to thank Johanna for her loving support during the entire thesis process.

Espoo, October 16, 2019

Kasper Kylmälä

Contents

Chapter 1

Introduction

The introduction will set the stage and background for the study. The background information of this study will be explained first, after which the study's partner organisation Startup Refugees is introduced to help understanding the environment in which this study was conducted. Finally the purpose of the study is stated.

1.1 Background

In the summer of 2015, the European Union began experiencing an unseen amount of people applying for asylum, a scale of which had not been seen since World War II. European countries were unprepared for the amount of refugees and the phenomenon was soon named the European refugee crisis. Finland was also affected, the yearly amounts of applications increased roughly 890 % compared to the previous year. The public sector was overwhelmed, new reception centers were established and the old centers worked over their capacities to house the arriving asylum seekers. The processing of asylum applications was congested all around Europe and reception centers became camps where one lives in a limbo between two states.

The fleeing refugees caught the attention of the world for a while, however the attention soon shifted towards the host nations and their abilities to integrate the newcomers as part of their societies. To live as part of a new society after experiencing radical, sometimes violent changes in one's life is not trivial by any means. The European nations have handled the amounts of arriving refugees with different strategies, some of which have been more successful than others in terms of integration. Some countries have experienced a rapid start in the integration process, while some have experienced

the refugees being marginalised. However, in the time scale of big historical events the past 4 years from the beginning of the refugee crisis is a relatively short time. There are still people in refugee camps waiting for their applications to be processed, waiting to start their lives anew. One could also argue that in the European countries there has not been enough time for the integration process to even actualise yet. The integration of refugees has raised a wide range of opinions within the host nations. Some political parties have fixed their attention in radically limiting the acceptance of refugees into the country, while some individuals and organisations have begun to speak out for novel integration models for a more efficient integration process.

There is a certain element that has gotten the attention of both refugee organisations and economic researchers: the employment of refugees. A job has been found to significantly ease the integration process, since it enables one to live an independent life and find a place to belong in a new country. Having the refugees employed also makes sense in terms of economics. To support unemployed refugees is a huge cost to the government and having a job significantly increases the refugees' standard of living. However, the employment rates and earnings of refugees are consistently lower compared to the native population everywhere in the world. This has raised questions among refugee organisations, politicians, and academics: why is it so hard to get refugees employed, could it be made easier? So far only hypotheses and theoretical speculation have been used to answer this question and there is a severe lack of evidence about the topic.

1.2 Startup Refugees

Startup Refugees (SUR) is a Helsinki based organisation with the main goal of supporting employment of refugees and asylum seekers. The organisation was established during 2015 as a reaction to the refugee crisis, to support arriving refugees in a new way based on employment support. To facilitate the integration process, to ease the economic burden of the government, and to simply get more people employed SUR connects refugees with employers in need of workers and also organises various educational opportunities. For more entrepreneurial refugees SUR also offers a free courses about entrepreneurship in Finland. Even though SUR's focus is on refugees and asylum seekers, the services are for immigrants to use as well.

As of October 2019 SUR had collected the data of over 3500 immigrants, of which most were asylum seekers. During the roughly four years they have

mediated over 700 jobs. While at it, SUR's employees noticed that this data of registered refugees and asylum seekers could be used for research purposes, which was the inspiration for this study. This opportunity is extremely valuable, since it makes it possible to conduct quantitative research about refugees, asylum seekers, and their employment with a relatively big data collection, to provide the highly needed evidence about refugee employment.

1.2.1 Employment process

The SUR employment process begins by interviewing refugees and asylum seekers about their background information, skills, education, and work experience. This is done by filling out a digital profile in a SUR's tool called Match, which is publicly available for use: `match.startuprefugees.com`. After registration, the refugees, asylum seekers, and immigrants are part of the SUR employment program, ready to be informed about a possible job offering by SUR's partner companies. As a bonus, Match also hands the refugees an automatically parsed CV in pdf format, designed for Finnish job markets.

When the time is right, partner companies inform SUR about their need of workers: how many and what kind of workers are needed. SUR's employees then search the Match database for suitable candidates. Before the job interview SUR arranges workshops for the candidates to teach them the basics of Finnish work life. The job interview process is covered in detail so everyone knows what is to come. This is necessary since for some this is the first formal job interview or their first time applying for a job in their life.

If the candidates are required to have any formal certificate for the job (hygiene certificate, alcohol passport etc.), SUR supports the refugees in acquiring them. The employing company then arranges job interviews and SUR's employees are there as well for support if any translation or general advice is needed. The company then decides which candidates are hired. SUR also looks after the persons to be hired, so that the job contracts are legitimate, the arrangements are legal, and the refugees are not being scammed.

1.3 Purpose of this study

This study's main purpose is to provide evidence to the question of why refugees tend to be worse off in terms of employment from the individual refugee's point of view, what makes it difficult to get employed as a refugee. While at it, it is also practical to look for factors that make it easier. This study exists to provide facts to the political discussion about refugees and

their integration in Finland. The political discussion around refugees tends to revolve around hypotheses, educated guesses, and beliefs, there is a dire need for statistical evidence. The wellbeing and even lives of tens of thousands of people revolve around the political decisions about refugees and asylum seekers in Finland alone and the weight of the topic is even heavier considering the whole of European Union. Scientific evidence regarding refugees, their integration, and employment can be considered critically important because of its scarcity and huge need.

Regarding SUR, this study provides a view into the results and impact of their first operating years. Even though the results are there to answer critically important questions about refugee employment, the statistical outlooks into the refugee profiles will provide ample information about the refugee population SUR has been serving and how SUR has been able to help them in terms of both quality and quantity.

Finnish researchers Sarvimäki, Joro, and Eronen have already had a look into this topic, providing valuable information about the economic situation of refugees in Finland, their backgrounds, the public employment services' effect on their employment, and their future plans and dreams. This research continues the work of Sarvimäki, Joro, and Eronen in its purpose of providing the Finnish policy makers with facts about refugees and asylum seekers in order to enable more informed decisions.

Chapter 2

Literature

This chapter begins by defining some important general concepts, in this case refugees, asylum seekers, and integration. After the definitions the basic economic mechanisms behind immigration are explored, leading to the arguments of why employing refugees is important.

2.1 Refugees and asylum seekers

The UNHCR (1967) defines refugees as people outside the country of their nationality or residence because of a well founded fear of persecution and are unable or unwilling to return to that country. This persecution can be because of race, religion, nationality, membership of a social group, or political opinions. The definition by the UNHCR (1967) makes refugees a special case of immigration, on whom a different set of laws apply. Refugees are required to comply with refugee specific laws and regulations of the country granting asylum in return for international protection. On the other hand the nation granting asylum recognises the rights of an asylum seeker, including but not limited to work, education, shelter, healthcare, and other rights comparable to other immigrants (Sisäasiainministeriö, 2011).

In practice the legal framework agreed by the (UNHCR, 1967) has more steps to it. The country granting asylum must first investigate whether the asylum seeker is in fact a refugee according to the definition by UNHCR (1967). Legal procedures take time, and the rate of asylum applications can vary heavily (Finnish Immigration Service, 2019d). Therefore a legal category of asylum seekers also exists (Ministry of the Interior - Finland, 2019). An asylum seeker is a person waiting for the decision about his refugee status and this can take years, making asylum seekers a relevant segment of

the immigrant population of a country (Finnish Immigration Service, 2019d; Hainmueller et al., 2016; Haverinen, 2018). An asylum seeker's application can be accepted and the asylum seeker legally becomes a refugee (Ministry of the Interior - Finland, 2019). If the application is denied, the asylum seeker is typically sent back to his or her country of origin (Ministry of the Interior - Finland, 2019). In practice the deportation process is a complicated one and the nation of origin might not accept the deportation, be able to receive refugees back, or in the worst case the nation of origin might not even exist anymore (e.g. the Soviet Union), this might lead to a situation where an asylum seeker is taken into custody. The asylum seeker has the legal right to appeal to the court of law for an unfair decision (Saarela, 2017; Finnish Immigration Service, 2019c).

During the refugee application processes in an EU country asylum seekers typically have diminished rights compared to refugees, they mostly stay in the reception centers, are under threat of deportation or moving, and are strangers to their new country of residence (Yijälä and Nyman, 2017). In this state asylum seekers are a vulnerable segment and live in a state of limbo, where they cannot go back to their home country, but the nation granting them refuge is taking its time to let them in (Yijälä and Nyman, 2017; Hainmueller et al., 2016). Meanwhile they are unfamiliar with the new country, local habits, and language, while having usually witnessed traumatic experiences in the recent past (Berry, 1992; Yle, 2019; ja hyvinvoinnin laitos, 2019). Typically in Europe the life as an asylum seeker is one of uncertainty, worries, restrictions, and boredom, which Haverinen (2018) calls "forced idleness" (from Finnish: "Pakotettu toimettomuus"). Depending on the refugee situation and the country of refuge, the life as an asylum seeker could of course be much worse, for example in a refugee camp (UNHCR, 2019; Smith-Spark, 2015).

2.1.1 Refugees and asylum seekers in Finland

After being accepted as a refugee in Finland, one becomes a special kind of immigrant, allowed to live one's own life with rights comparable to other immigrants, no longer under the threat of deportation. A refugee is still treated as a foreign citizen though (Finnish Immigration Service, 2019a). An asylum seeker's situation is different. An asylum seeker is constantly under the threat of deportation, he cannot legally leave the country and his rights to social security and work, among other things, are restricted (Sisäasiainministeriö, 2011). Both asylum seekers and temporary refugees are able to live

in a reception center, but can live in a private apartment if they are able and willing (Sisäasiainministeriö, 2011). This is usually not the case since asylum seekers tend not to have significant sources of income of their own (Sarvimäki, 2017; Yijälä and Nyman, 2017). A reception center is required by law to provide decent accommodation, food, education, hobby activities, social services, healthcare, and work opportunities. However Haverinen (2018) claims that this only works on paper, in practice reception centers focus only on basic necessities and tend to function only as a storage for refugees.

The Finnish legislation grants an asylum seeker the right to work if he has stayed in Finland for three or six months, depending on whether the person has a valid ID, but he is not allowed to start a business (Finnish Immigration Service, 2019b). This is decent compared to other European nations, for example in Sweden and Norway asylum seekers can work immediately with a valid ID, but in Ireland they are not allowed to work at all (Legrain, 2017). The right to work needs to be checked by the employer from the Finnish Immigration Service, who keeps track of the status of aliens in Finland (Sisäasiainministeriö, 2011). If the right to work leads to long term employment, there is a possibility for the asylum seeker to apply for a residence permit based on work, which could lead to one's status being elevated to an immigrant, which usually serves the same purpose from the asylum seeker's point of view as being accepted as a refugee (Finnish Immigration Service, 2019e). Finland is compelled by the UNHCR (1967) and European Court of Human Rights (1971) to maintain the wellbeing of refugees and asylum seekers and has made clear that its goal is to facilitate their successful integration into the Finnish society (Valtioneuvosto, 2015; Ministry of the Interior - Finland, 2019).

For simplicity, this study will use the term refugees to refer to both refugees and asylum seekers from now on. If the two groups are both mentioned, then refugees will refer only to refugees, asylum seekers will be referred to separately in more specific situations.

Around 81 000 foreigners applied to live in Finland during 2018, roughly 4 500 of these were asylum seekers (Finnish Immigration Service, 2019f). The number of asylum seekers peaked during the year 2015 as part of the European refugee crisis, that year around 32 500 people applied for asylum in Finland, a 890 % increase compared to the previous year (Smith-Spark, 2015; Finnish Immigration Service, 2019f). Even though Finland has its history with Vietnamese and Somali refugees among others, the amount of people arriving from these countries is small compared to the amount that arrived during and after 2015 (Hangartner and Sarvimäki, 2016). Clearly the largest nationality

among the Finnish refugee population are Iraqis and it is possible that they already form the majority of refugees in Finland (Sarvimäki, 2017). According to an overview by UNHCR (2019) of the whole population of refugees and asylum seekers 19 % are women, 30 % children, and 51 % men as of 2017. The entire Finnish population of refugees, asylum seekers, and stateless people was roughly 27 000 in 2018 (UNHCR, 2019). In her study Joro (2019) estimated that refugees in Finland tend to be fairly young, majority of them being under 35 years old. The data from Hiekkavuo (2016) also supports this, reminding that the overall immigrant population is significantly younger than Finland's native population. From the education and profession point of view Joro (2019) pointed out that over 25 % of refugees and asylum seekers reported having at least studied for higher education, but also mentions that this number must taken with a grain of salt and is probably lower in reality. Work experiences of the same population were dominated by labour intensive jobs, like repairs, construction, sales, and artisanry (Joro, 2019).

2.1.2 Refugee acculturation and integration

Berry (1992) defined acculturation as a process of change that leads to different adaptation outcomes when two different cultural groups come into contact in a society. When immigrants or refugees arrive to a new country with a different culture, various kinds of changes happen on the individual level, making the adaptation an outcome dependant of various internal and external factors (Berry, 1992). Some of these factors are based on the attitudes of the immigrants themselves, Berry (1992) calls these acculturation strategies. Acculturation strategies are a function of one's attitudes on two questions. The first being whether one should retain one's cultural identity. The other is about one's willingness to interact with the foreign culture. The attitudes towards these two questions, whether negative or positive, can then be divided into four acculturation options which are then both strategies and outcomes: Marginalization, Separation, Assimilation, and Integration. See figure 2.1 below for a visual model.

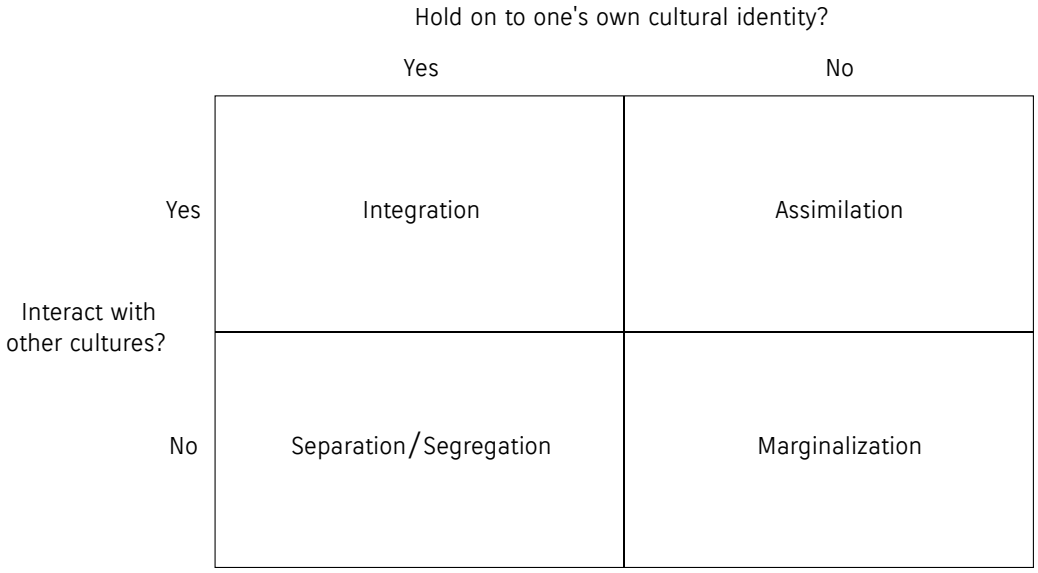

Figure 2.1: The four acculturation strategies

Of all the four options *Marginalization* is the most complex. It is the combination of answering both questions with a strict "no", swaying away from one's cultural identity, while refusing to interact with the foreign culture (Berry, 1992). From the acculturing individual's point of view it is characterized as rebelling against the dominant society, while simultaneously going through feelings of alienation, loss of identity, and serious acculturative stress (Berry, 1992). This can be caused by either exclusion or withdrawal, but it is typically a situation where neither the dominant culture or the acculturing individuals benefit (Berry, 1992).

When one's answers to the questions are to hold on to one's own cultural identity, while rejecting the larger society, Berry (1992) calls it *Separation*. However, *Separation* also includes some complications, since there is a significant difference depending on which culture group (dominant or non-dominant) controls the acculturation process (Berry, 1992). When the dominant culture has the upper hand and their agenda is to "put people in their place" it is called *Segregation*. On the other hand if the non-dominant group controls the situation and withdraws from the dominant culture the option is called *Separation* (Berry, 1992).

Assimilation happens when the acculturing people abandon their own cultural identity and get absorbed into the larger society or when many groups merge into a single completely new culture (Berry, 1992). A classic example

of this is the melting pot concept which has been used to portray the United States among other multicultural environments (Berry, 1992).

Finally, *Integration* is defined as willingness to interact and co-exist with the dominant culture, while holding on to one's own cultural identity (Berry, 1992). When the Integration option is widely adopted, various ethnic groups exist and co-operate in a larger society (Berry, 1992). Of all acculturation options *Integration* has been most preferred, since it serves the larger society while being best for the well being of the acculturing immigrants, which is found to be extremely important in the case of refugees (Yijälä and Nyman, 2017; Yle, 2019; ja hyvinvoinnin laitos, 2019). As already mentioned in the subsection 2.1.1, *Integration* is also the most preferred outcome of the refugee acculturation in the eyes of the Finnish government (Valtioneuvosto, 2015). However, the topic of immigration and refugees is not without its controversy and a decent amount of anti-immigration and anti-muslim political movements have gained ground during and after the refugee crisis (Hangartner and Sarvimäki, 2016; Timsit, 2017). Immigration is therefore a topic that keeps dividing people even within the host countries themselves, which does not make the much needed integration processes easier, but could instead lead to Segregation (Hangartner and Sarvimäki, 2016; Berry, 1992; Timsit, 2017).

2.2 Economics of immigration

It is not uncommon to see immigrants portrayed only as economic burdens, since they need to be taken care of and they generally do not have a way of earning a living by themselves (Legrain, 2016). This view is not completely false, since it is true that taking care and integrating immigrants costs money, takes time, and is not an easy task (Hangartner and Sarvimäki, 2016; Borjas, 1994). However a commonly overlooked argument outspoken by Legrain (2016): even though refugees do not migrate primarily for economic purposes they can and will, if properly managed, be an economic benefit for the country of refuge like other kinds of immigrants (Legrain, 2016). Even though Legrain (2016) talks about refugees, refugees are a special kind of immigrants, so it makes sense to apply it to immigrants in general as well, this view is also supported by Borjas (1994, 2013).

Immigration in general has its usefulness in increasing and diversifying the host country's labour force (Sarvimäki, 2010a; Borjas, 1994). On the downside immigrants typically have an economic disadvantage, because they are strangers in a new, foreign environment (Borjas, 1994). This in turn

means that employment rates and earnings of new immigrants are consistently lower than that of the natives (Sarvimäki, 2010a). This economic situation also explains why the recently immigrated people tend to use more of the welfare state funds compared to the native population (Sarvimäki, 2010a; Borjas, 1994). Thus in the beginning of immigration, the supply of labour increases while the state needs to spend more money supporting outsiders. There is a clear and obvious consensus among researchers that immigrants themselves benefit from the immigration, as well as some parts of the society because of the increase in labour force (Sarvimäki, 2010b; Borjas, 2013). However, who and how many of the native population suffer from immigration is a topic of debate (Sarvimäki, 2010b; Borjas, 2013).

An analysis based purely on basic labour economics would suggest that in the short term the increase in labour force would cut the wages and employment of the native population. In the neutral case where the immigrants would have an identical education profile to the native population this change would be insignificant in practice, since the labour pressure would be distributed evenly among the entire work force (Sarvimäki, 2010a). However, a problem emerges in a situation where the immigrant population's average education is lower than that of the native population, which is a completely realistic case since education levels are not evenly distributed around the world. In this case the increase in labour supply cuts the wages and employment of the lower educated segment of the population, which is typically something the native population does not welcome with open arms and can lead to growing anti-immigrant attitudes (Sarvimäki, 2010a; Peri and Sparber, 2008). More recent and complex analyses suggest a different outcome though. Sarvimäki (2010b), Peri and Sparber (2008), and Legrain (2016) all suggest that the addition of immigrants to the labour force gives the native population, which is more prepared for tasks requiring more sophisticated coordination and communication, a chance to specialise in jobs requiring those skills, rather than competing for the same jobs with the immigrants. They also suggest that the state spending on supporting immigrants might stimulate the economy, since employed immigrants would be both producing and consuming like any other resident (Sarvimäki, 2010b; Peri and Sparber, 2008; Legrain, 2016). With this in mind, the negative effects of immigration to the low educated part of the society could be relatively small, and the effect of immigration to the society in general could be net positive (Borjas, 2013).

If we take the long term perspective, immigration can be seen in even more positive light. Chiswick (1978), Borjas (1994), and later Sarvimäki (2011) claim that the economic situation of immigrants gets better over

time. With time immigrants learn the habits of the new environment, form social connections, get employed, get further on their career, learn the local language, integrate better into society, and gain overall experience in life (Sarvimäki, 2011). On average, employment rates and wages in immigrant populations rise when immigrants spend more time in their new home, because of this the state's social security spending on them decreases over time and can then be allocated somewhere else (Sarvimäki, 2011; Borjas, 2013). This makes even more sense if the probability of integrating or assimilating into the society is higher, in contrast to a situation where massive separation, segregation, or marginalisation occurs systematically (Chiswick, 1978; Berry, 1992; Sarvimäki, 2011). In his study Chiswick (1978) argued that immigrants in the USA were even able to surpass native population in earnings after 10 - 15 years, however Sarvimäki (2011) was not able to witness the same results in Finland. Given enough time, the studies of Borjas (1994) and Sarvimäki (2011) support the claim by Legrain (2016) that immigrants can be a real economic benefit for the host nation.

2.2.1 Economics of refugees

Economically refugees are very similar to immigrants. Legrain (2016) compares the economic case of refugees to immigrants with an initial investment that needs to be payed before the economic benefits roll in. However, Legrain (2016) points out that compared to the long term benefits, the initial investment is relatively small and it comes with the added utilitarian ethical bonus for actually saving numerous human lives from serious danger and violence (Mill, 1879).

Legrain (2016) clearly points out the critical upfront costs for a successful refugee integration that he considers the initial investment for refugees to be able to become a net benefit for the host nation. These costs include food, clothing, shelter, basic income support, healthcare, help with mental trauma, language training, schooling, and the necessary administrative costs. Legrain (2016) also points out that the initial investment gets smaller the less time refugees spend in the hands of public assistance. If the host nation can get refugees into work sooner rather than later it would save money, give the refugees a chance to start contributing to the economy, and would facilitate integration (Legrain, 2016). Legrain (2016) and Hangartner and Sarvimäki (2016) approximate that with the policies of 2016 one refugee in a EU country costs around 10 000 - 12 000 € for the host country as an initial investment. Hangartner and Sarvimäki (2016) remind us that the accurate estimation

of the long term costs of fully integrating high amounts of new residents is difficult, but Legrain (2016) and Hangartner and Sarvimäki (2016) point out that given the fertile ground for successful integration and the possibility to work will cause refugees to be a net benefit for the host society. The economic benefit is at its highest if the refugees arrive at the optimal work life age of 25 - 30 years old (Hangartner and Sarvimäki, 2016). Because of the ongoing political schisms in various nations about immigration and refugees, the topic of refugees' economic worth has become an even more valuable argument, since humanitarian reasons alone might not be enough (Timsit, 2017).

2.2.2 Refugee employment

As mentioned in the last subsection 2.2.1, getting refugees to work is a critical element for the economic benefit of the host country. In addition, it has been widely accepted that employment is also an important method for integrating refugees into the society and for the refugees' wellbeing, while at the same time it is clear that refugees struggle to find employment in their new countries of residence (Aycan and Berry, 1996; Feeney, 2000; Krahn et al., 2007; Legrain, 2016, 2017; Haverinen, 2018). Hangartner and Sarvimäki (2016), Yijälä (2014), and Hainmueller et al. (2016) have also found out that longer asylum processing makes employment even harder. Since a number of researchers and organisations call for the employment of refugees rather sooner than later, it can be argued that we are currently witnessing the birth of a new idea that I will call: *the rapid integration of refugees*.

For all people work can give something meaningful to do, it gives us a chance to earn our own living, to learn and improve, to meet new friends or partners, we work for status, wealth, and our place in the world (Kurzgesagt - In a Nutshell, 2017). For refugees work means something even more, as Legrain (2017) says: "As well as being good for society, working benefits refugees themselves - and it is what they want to do. While they have suffered greatly, they typically do not want to be forever treated as victims or charity cases. They want to start rebuilding their lives and become self-reliant again. In addition to generating income, work makes refugees feel valued and proud to be giving something back to the society that has welcomed them. When asked "what makes you feel integrated?" most refugees respond: "to have a job.""

Even though researchers agree that work is a critical part of the integration process and a lot of policy proposals have been made to facilitate refugee employment, various kinds of barriers exist for refugees to find stable

employment. These barriers include laws restricting refugees' rights to work and entrepreneurship, risk of deportation or forced moving, mismatch of skills required for the local job markets, inadequate language skills, lack of clear paths to employment, lack of advice towards employment, discrimination, strict labour market regulations, reception centers' impractical locations far from job opportunities, long asylum application processes, inadequate education among refugees (illiteracy for example), lack of local work experience, lack of work experience in general, and reluctance to accept foreign educational credentials by host nations' employers and officials (Shields and Price, 1999; Krahn et al., 2007; Isphording et al., 2014; Legrain, 2016, 2017; Haverinen, 2018). In addition to the barriers for employment, a wide gender gap also exists among refugees - women are significantly less likely to get employed than men (Legrain, 2017). Krahn et al. (2007) remind us that these barriers affect all refugees, including the highly educated, by increasing the probability that highly educated refugees will only find employment in low education jobs. The effect of long asylum application processes as a barrier for employment has been validated in various studies (Hainmueller et al., 2016; Desiderio, 2016; Legrain, 2016, 2017; Hangartner and Sarvimäki, 2016). Several factors have clearly been recognised as plausible barriers but consistent knowledge of their existence, severity, and order of importance is lacking. A barrier that has been validated and is commonly agreed upon in the literature is the lack of native language skill (Shields and Price, 1999; Isphording et al., 2014).

An interesting paradox exists regarding the education barrier, many researchers have found out that refugees arriving to OECD countries tend to have lower level of education compared to the native population (Aycan and Berry, 1996; Eronen et al., 2014; Desiderio, 2016; Joro, 2019). At the same time Krahn et al. (2007), Bevelander and Lundh (2007), and Joro (2019) have found that a lot of employed refugees are clearly overqualified for the job in terms of education. This paradox implies that other barriers than education are also active, holding refugees in place by downward mobility (Krahn et al., 2007). Significant downward mobility combined with the desperation that refugees experience might explain why refugees tend to get employed in low paying jobs, usually of manual labour - a situation of this kind seems to encourage an attitude where any kind of job will do (Legrain, 2017; Joro, 2019; Siironen, 2019).

In the most recent Finnish study on this topic, Joro (2019) contributed to the research by conducting a combination of an explorative study into the backgrounds, skills, and experience of the refugees in Finland in addition to

a qualitative study into the details of the experiences and future plans of the same persons. This very recent research by Joro (2019) shares many similarities with this study, including its purpose and sample population, therefore it will work as an important comparison for this study's results. Even though not as recent, the works of Eronen et al. (2014) and Sarvimäki (2017) have been also studying the employment and job market integration of refugees in Finland, so their findings are also definitely used as both comparisons and inspiration.

Chapter 3

Research questions, goals, and purpose

As explained in the subsection 2.2.2, the rapid employment of refugees is critically important in their integration to new societies, but accurate knowledge about the barriers inhibiting refugee employment is lacking. A wide array of possible barriers has been identified in both Finnish and international environments, but only a few studies have been conducted to validate their actual effect in different populations and environments. Therefore this study contributes to existing research by testing which factors within refugees themselves act as barriers that inhibit their employment in the Finnish job markets. As a bonus this study provides detailed information the refugee population in Finland.

The goal of this study is to identify which factors within refugees themselves constitute barriers for their employment in Finland, which can be divided into two research questions.

1. Which refugee background factors have a direct connection with employment?

2. Which refugee background factors have an indirect connection with employment?

The practical implementation of the study will also be supported by some statistical outlooks, to figure out what kind of refugees there are in terms of their backgrounds. This statistical outlook will also help in getting a deeper understanding on the topic, before digging deeper into the connections between the factors. In this study, backgrounds are defined as general information about individuals that might be relevant for employment, these include

nationality, gender, age, profession, native language, Finnish language skills, English language skills, entrepreneurial plans, and work experience. To clarify - external factors that might affect the employment of refugees, e.g. government subsidies/regulation on labour, economic factors such as the availability of jobs in general, or the foreign language proficiency of Finnish employers, are outside the scope of this study. This study focuses on the differences between individual refugees and the possible factors that make Finnish employers favor some of them while ignoring others.

Since barriers for employment can be either factors or their lack of, it is reasonable to look for both positive and negative connections between the factors and employment. Later it might be concluded that this factor, or the lack of it, could be considered a barrier for employment. This is why barriers for employment are not explicitly mentioned in the research question.

Chapter 4

Methodology

This chapter explains the methods used in the study and the practical implementation in detail. First the choices behind the methods are explored and hypotheses are presented, later on data collection method is explained, followed by the exploration of the various steps of the data-analysis and logistic regression models, last the plan for conducting surveys is explained.

4.1 Research methods and hypotheses

For easier readability the research questions are again presented below, for the reasoning behind them, see chapter 3.

1. Which refugee background factors have a direct connection with employment?

2. Which refugee background factors have an indirect connection with employment?

4.1.1 Methods and research plan

Given the nature of the research questions and the kind of data available from SUR, this study could be implemented both from a quantitative or a qualitative perspective (Easterbrook et al., 2008). There is plenty of qualitative data about people, their skills, education, personality, and past work experiences in the SUR database, although a lot of quantitative and quantifiable data exists as well. Both perspectives could have been able to figure out barriers for refugee employment, however with different research outputs (Easterbrook et al., 2008).

At its current state, the literature regarding the refugees and their employment contains plenty of both qualitative and quantitative information. However, the barriers for refugee employment have been explored only from the qualitative perspective and most of the explanations behind the phenomenon are purely theoretical. Therefore one could argue that there is a serious lack of hard evidence about the topic, which a study with a quantitative approach with the SUR's data could be able to provide (Easterbrook et al., 2008; Schutt and O'Neil, 2013). Given this unique opportunity of having large amounts of quantitative data from SUR, the lack of evidence about refugees' barriers for employment, and my adequate personal capabilities and interest in data science, this study will be that of a quantitative data analysis.

To support the results of the quantitative data analyses a small qualitative study is used in triangulation to validate the results answering the research questions (Jick, 1979). Triangulation in general is used to bolster the validity of a research or its findings, as Miles & Huberman said in Mathison (1988): "...triangulation is supposed to support a finding by showing that independent measures of it agree with it or, at least, don't contradict it". In this study the qualitative results are used to check whether the results from the quantitative part actually make sense. It will also be used for shedding light on new points of view about the barriers for refugee employment, something that either the data collection by SUR or the analysis of this study might have missed (Jick, 1979; Mathison, 1988).

The plan for implementing the study is to start with the statistical outlooks. The statistics will be compiled to examine the entire sample population and its characteristics, but also to examine only the refugees who have gotten employed. Having the two distinct sets of descriptive statistics makes it possible to make high-level comparisons between the populations and for one to have an overall understanding of the SUR's client refugees in general. The research question 1 will be answered by creating a logistic regression model to explore which factors have significant connections to refugee employment through SUR. The research question 2 will be answered by creating more logistic regression models between the independent variables, to figure out whether some can be connected to employment indirectly through other factors. In addition, surveys will be conducted to discover any kind of factors that might be affecting refugee employment in the entire SUR employment process.

4.1.2 Hypotheses

The research questions are of diagnostic nature, therefore having hypotheses to be tested is reasonable. Regarding the possible factors that could be analysed from the SUR data, the null hypothesis for the research questions would be: no factor has any connection with either employment or other factors, which would imply there not being any barriers for employment.

Credible evidence exists that the lack of native language skills is a serious barrier for employment (see subsection 2.2.2) and detailed data about refugees' language skills is abundant in the SUR database, thus it would make sense to hypothesise that lack of Finnish language skill has a direct connection with lower employment. Education is also a factor widely discussed among researchers, but it remains unclear whether it acts as a barrier for employment, whether it has an indirect effect on employment through some other barrier, or whether it has any significant effect at all. Regarding the indirect connections, even within the Finnish native population it has been proven that women have slightly lower rates of employment compared to men, which applies to refugees as well (Legrain, 2017; Tilastokeskus, 2018). This implies that gender has either a direct or an indirect connection with refugee employment. The literature does not emphasise gender as a serious barrier for employment though, therefore it is reasonable to hypothesise that the connection between gender and employment is indirect. However, it remains unknown through which factor gender would have its indirect connection with refugee employment.

Based on the null hypothesis and the arguments presented in the previous paragraph, the hypotheses for the research questions are formulated as:

Hypothesis 1: Only Finnish language skills have a direct connection with refugee employment.

Hypothesis 2: Only gender has an indirect connection with refugee employment.

4.2 Data collection

As mentioned in the subsection 1.2.1, the SUR data is collected by filling Match profiles with the refugees wanting to find a job. There are two ways of how the registration is done. The simplest way for a person in need of a job is to head to the Match website, create an account, log in, fill the profile with one's information, and hope for a call from SUR about a suitable job

opportunity. However, most of the profiles are collected in events called Match clinics. Match clinics are organized in either public workspaces, reception centers, or hosted by a SUR partner organisation. In Match clinics SUR's employees and volunteers work together with refugees to fill the profiles, since Match is not yet user friendly enough to be filled alone without any trouble and is only available in English. Translation help is often required since most of the refugees do not speak English or Finnish. Regarding the information of each profile, there are a lot of questions to be answered which form the basis for the data of this study. Most of the questions are listed here in a simplified format, with the answer formats included.

1. Personal information
    (a) Date of birth [Date]
    (b) Nationality [Country selection]
    (c) Gender [Male, female, other]
    (d) Place of residence [Free text]
    (e) Do you have a passport? [Yes/no]
    (f) Do you have a residence permit? [Yes/no]
    (g) Date of arrival to Finland [Date]
    (h) Migri-ID (this is a unique identifier used by the Finnish immigration agency) [Number]
2. Language skills
    (a) How well do you know Finnish?
        - Scale from 0 to 5, from none to mother tongue
    (b) How well do you know Swedish?
        - Scale from 0 to 5, from none to mother tongue
    (c) How well do you know English?
        - Scale from 0 to 5, from none to mother tongue
    (d) What other languages do you know and how well?
        - Scale from 0 to 5, from none to mother tongue
3. Education
    (a) What is your highest level of education?

- Scale from 0 to 6, numbers representing education levels in the following order: No education, pre-school, elementary school, upper secondary school, vocational school, university degree, doctoral degree

(b) Topic of studies [Free text]

(c) Start year of studies [Number]

(d) End year of studies [Number]

4. Work experience

(a) Your professional title [Free text]

(b) List all your previous work experiences

- Job description [Free text]
- Company name [Free text]
- Starting date [Date]
- End date [Date]

5. Skills

(a) Select all your skills from a huge list of various skills

- Rank every selected skill in a scale from 0 to 5

6. Miscellaneous questions

(a) Interests and hobbies [Free text]

(b) How would you introduce yourself to a future employer? [Free text]

(c) Do you want to establish a company? [Yes/no]

(d) Do you have a business idea? [Yes/no]

(e) Are you interested in something besides work? [Selection]

(f) Do you have any Finnish work certificates? [Yes/no for all Finnish certificates]

(g) Do you have any other work certificates? [Free text]

As one might notice, there are a lot of possible problems in this way of collecting data, which should be taken into account when interpreting the results of this study. First of all the translation might cause problems with

understanding some parts of the interviews, thus paving the way for a faulty answer. The question about education tends to be a problematic one, since the answer scale is based on the Finnish education system and its levels, there have been numerous problems trying to transpose the accomplishments in a foreign education system to the Finnish scale. Combined with the problems of translation, the education question is one of the hardest to get reliable answers to. SUR also has no way of actually verifying whether the person filling the profile actually gives truthful answers, this combined with the fact that the refugees have a chance to get a job incentifies them to exaggerate their answers. However SUR's employees do their best and usually slightly test the refugee, to maximise chances for a truthful answer. A lot of trouble in the data analysis will probably be caused by the profile form having a lot of free text fields, this will provide answers without any standard form, thus being more difficult to quantify. The free text fields also give ample room for a common human mistake: typos. Both typos and the non-standard answers can together force corrections, speculation, guessing, or deletion of data by the researcher, which can cause biases in the results and having less data to analyse. Another simple problem arising from this method of data collection is that even though people have a chance to update their profile, they tend not to, thus the data saved in the Match database is typically not up to date. Also, there is a single question which seems to be misused by the refugees as well as SUR's employees (because of a lack of a better field): the "Professional title" field is typically used as a field where one lists everything one is able to do and in some cases the kinds of jobs one would like to apply, thus the content is not consistent and everyone tends to have their own interpretation about how it should be used. Finally the problem that has caused problems even in SUR's day-to-day operations: a lot of people have multiple profiles in the Match system, some of them even have slightly different information. The detection of the duplicate accounts is typically not a trivial task.

In addition to the Match system, Startup Refugees has kept records of most of the jobs they have mediated. The employing company, job description, starting date, city, and contract type were all saved, and in some cases even the person's name and Migri-ID. Data of 593 employments was available, which one has to remember, is different from the amount of people employed.

Even though the data collection has its shortcomings, data covering the work life skills and experiences of a large number of refugees is something that is fairly unique in refugee studies. This kind of data is so rare, since very few organizations have the resources or incentives to gather it and govern-

ments only tend to compile statistics about asylum applications and decisions. Therefore analysing this collection of data has its chances to yield novel results which could be useful for both policy makers and for future research. This kind of data collection method makes this study not one of "gathering and analysing specific data according to the experiment design", but one of "making the most of existing data of limited quality".

4.3 Data analysis

This section walks through the journey of analysing the data collected by SUR to answer the research questions. The section uses the widely adopted guidelines of Schutt and O'Neil (2013) to have a systematic structure for the whole data analysis process. The actual methods used that provide results are basic statistical calculations and logistic regression modeling.

4.3.1 Data requirements and selection

Since the data collected by Startup Refugees was collected for the purpose of employing people, some changes need to be made for the data set to be suitable for the logistic regression analysis. First requirement for the data is that it should be organised as a dataset with individual persons represented as rows, for this reason duplicate accounts should be deleted. Every piece of data should be in a quantifiable format of any scale of measurement: nominal, ordinal, interval, or ratio (Stevens, 1946). The data should also be at least somewhat up to date, information that references the past rather than the present is more reliable because information about the present situation tends to constantly change with time, while information about the past remains constant. Finally because of the nature of logistic regression the amount of independent variables is limited, since this will cause either data sparseness, multicollinearity, or separation, which will all produce unreliable results (Concato et al., 1996). In the context of this study, too many predictors will probably be caused by either using ordinal predictors with too many possible states, or too many predictors in general (see subsection 4.3.4 for further details).

With the requirements in mind, a dataset is to be created with the data collected from the Match database. The dataset will consist of rows representing individual persons, with multiple columns representing the individual variables and one column representing the dependent variable. The dataset

will be handled with WPS Spreadsheet software for convenient calculation and data processing.

Starting from the top of the Match profile questions (see list 4.2), the first data selection decision is to discard the data referring to present, this means that the data about one's place of residence and passport will be ignored in the analysis. Both data about the refugees' passport and place of residence have a high probability to change in a short amount of time, therefore it can be considered unreliable.

The date of arrival would be an important factor to be taken into account, at least according to the literature. However, the arrival date column has a massive amount of empty entries, since it is not a required field. Most of these empty entries are caused by either refugees or immigrants, since the question is aimed towards asylum seekers. The problem gets more complicated because there is no data to distinguish between immigrants, refugees, asylum seekers, and persons who were asylum seekers at the time of Match registration who have then become refugees. This is why the column of arrival dates is discarded.

Migri ID will not be used for the analysis itself, but it will be useful for looking for duplicate accounts in the dataset, and pairing the data from the employment records with the data from Match, so the data from the correct employment will be combined with the right person's data.

Language skills and education level will be measured by the scale used by SUR as it is, but the amount of languages will be restricted to the three most significant in Finland: one's own native language, one's skill of Finnish, and one's skill of English. Work experiences will be measured as the total number of jobs the person has had. This is done to quantify the information about work experiences by assuming that persons with a higher number of work experiences tend to have more work experience in total. The problem with the work experience data is that both starting year and end year were asked as well, but were not required, thus there is no consistent way of measuring the total amount of work experience one has in years.

Because of the limit of predictors one can use in logistic regression, the skills will also be discarded from the analysis. This huge amount of skills available in Match compared to the amount of data of persons employed would cause notable data sparsity. Also a large proportion of the work life skills one has can be explained by one's profession, education, and work experience.

Some of the miscellaneous questions will also be ignored in the analysis. Interests and hobbies, and the introduction to a future employer are difficult to quantify. These two are probably not a significant factor in employment,

and these questions are usually the most confusing ones to people from outside western cultures. The work certificates on the other hand are valuable information in itself and easily quantifiable, but the information about them also refers to the present, thus the data is often outdated. SUR also supports the recruits in acquiring the necessary certificates to minimise the risk of it becoming a major barrier for employment, since it is easily avoidable. The binary data about the willingness to establish a company and having a business idea will be used in the dataset without any modifications. The database ID will be used as a control variable as it is, more information about it is available in subsection 4.3.2.

At this point, a sample of the dataset could look like the one below (figure 4.1), with eleven columns representing the variables for the analysis, this data set is only for illustration and does not represent the actual data. As one can see, there are anomalies in the birthdates, duplicate rows exist, and the profession column is not in a quantifiable form. Before any analysis is made, data processing and cleaning needs to be done. In this stage there are 3 263 persons in the dataset and data about 593 employments exists.

ID	migri-id	gender	birthdate	business	idea	mothertongue	finnish	english	nationality	education	profession	workexp
2	1122334	male	1992-12-17	0	1	Arabic	0	4	Iraq	5	Repairing cars	0
837	1514132	female	1953-01-01	0	0	Arabic	0	2	Iraq	4	dray clean warker	0
999	1798766	male	1990-02-01	1	0		2	0	Iraq	0	Journalis	4
2897	1798766	male	1990-02-01	0	0	Kurdish	0	1	Iraq	1		0
543	00998877	other	1970-06-30	1	1	Dari	1	0	Afghanistan	2	tailer	2
13	0998877	male	1970-06-30	0	0	Dari	2	3	Afghanistan	0	Tailor	1
2200		female	2025-03-28	0	0	Arabic	0	0	Jordan	1	Welder in aircraft	0
1447	5443322	male	1955-01-01	0	1	Somali	1	0	Somalia	3	Barber & Truck drive	3

Figure 4.1: A sample of the possible state of the data

4.3.2 Data processing and cleaning

The first data cleaning to be done is the removal of duplicate entries, this is done mainly with the help of Migri-ID. If a single Migri-ID can be noticed twice, it would imply a duplicate profile, and the information of the two rows will be merged. In a case where there would be contradictory data in some fields between the duplicate rows, the data from the newer one will be used. What makes the duplicate detection harder though is the matter of typos. There were a lot of cases where there were similar profiles in terms of data, but with slightly different Migri-ID, but also completely different profiles with the same Migri-ID. In both of the cases the personal information of the profiles was checked by using the Match tool's official admin view, where it was possible to search people by their Migri-ID and arrival dates to get the

name of the person. If the name of the persons having similar Migri-IDs was the same, the profiles were merged. If it was clear that the duplicate rows resemble one person, the profiles were merged. If the case was not clear, the merging was not done and no rows were merged. In total 114 rows were deleted due to merging, the amount rows in the dataset being now 3 149.

There are also some problems with birthdates. Since a lot of refugees do not know their birthdates, in many nations the official state processes for population statistics do not exist, the practical way to mark these in Match is to use the date January 1st of the said year. This creates unnatural biases in the date statistics, therefore only the year of each date will be used. However, there are also some dates which can be attributed to either typos or trolling, since some of the persons reported being born in the future. The profiles with their birth year in the future are only a handful, so they will be left as they are, but will not be taken into account when calculating statistics, see subsection 4.3.5 for further details.

There are also some empty entries in the nationality column, as well as a lot of empty entries in the mother tongue column. This will not be a problem while calculating the basic statistics, but it will be in the logistic regression. For this reason the cleaning will be postponed after the statistics have been compiled.

The professional skills one has are definitely critical in determining one's employment, this is why it is important to determine the most probable profession of every profiled person in a systematic way. For this, a list of professions is created which would cover every profiled person, but also would describe the actual profession, so the statistical results would be easy to understand. Because of the predictor limits of the logistic regression it is also probably necessary to group the professions into second order categories, so there is a way to reduce the amount of states of an ordinal variable. The choice of the profession for each profiled person is made based on three pieces of data: the person's "Professional title" answer, the title of one's highest level of education, and one's latest job description of the latest work experience. If all these are empty, the category of one's profession would become "None". If there are multiple possible professions applicable, the most specialised and the one requiring the highest level of education is chosen.

All professions to be used are based on case-by-case analysis of the profiles. For example, if there are police among the population, a profession of police will be made. Some generalisations will be made, such as all welders are marked as construction workers, all military officers are grouped together with soldiers, and herders and farmers are all generalized into a group called

agricultural workers. The professions are also grouped into second order categories. The grouping can be done in a number of ways, the decision was made with the help of SUR employees, to provide informative statistics from the point of view of refugee organisations and policy makers. The main criteria in the categorisation was the kind of education/specialisation required for the job, the exception being service work, under which are categorised both different kinds of therapists as well as cleaners. In the A.1 one can find a categorised list of all the 56 professions used, with explanations included. In essence, there are 11 second order categories: None, industry & agriculture, artisans, technicians, logistics, health & care, public officials, security, service work, specialists, and other.

The employment records also need cleaning and processing. The jobs of the employed ones are classified into the professions described in the previous paragraph. The employments are combined with rows in the main dataset with the help of the Migri-ID, this makes it possible to mark a 1 on every row representing a person who has been employed through SUR, while keeping other rows marked with a 0. Another column is dedicated to inform what job the employment was for, based on the profession classification, which will be used to compile statistics about the employments. Even though there is data from 594 employments, because of persons with multiple employments and employments without any data about the person employed, it is possible to combine the employment data to only 220 persons. However this is still enough to provide credible results through logistic regression. And now since duplicates have been found and person data has been combined with employment data, Migri-ID column is deleted.

There still exists an interesting variable which is actually a byproduct of the refugee registrations: the Match database ID. The database ID is an integer, which gives every person a unique identifier in the database, it is basically a number representing the order of people registering to Match, starting from 1. Even though the number seems arbitrary, the profiles that have existed in the Match system for a longer time, thus having a smaller ID number, naturally have had higher chances of getting employed, this is why the database ID is chosen as a control variable in the model. Also there are no database ID numbers missing, no typos, or duplicates, so there is no need for any data processing or cleaning, thus it will be used as it is.

With all necessary data processing and cleaning done, the data set is ready to be used to answer the research questions. Now it contains 13 columns, of which 12 are for independent variables: gender, birth year, willingness to establish a business, having a business idea, mother tongue, Finnish skill

level, English skill level, nationality, education level, profession category, work experience, and the control variable database ID. In addition there is one column for the dependent variable: having been employed through SUR before a database copy was created for this research (April 2019). An example of the final dataset would look like the one in figure 4.2. All the variables with their scales of measurement, how are they defined, and role in the study can be seen in the table 4.1 (Stevens, 1946).

Variable	**Scale**	**Type**	**Role**
Gender	Nominal	Binary	Independent
Birth year	Interval	Discrete	Independent
Business	Nominal	Binary	Independent
Business idea	Nominal	Binary	Independent
Mother tongue	Nominal	Categorical	Independent
Finnish	Ordinal	Discrete	Independent
English	Ordinal	Discrete	Independent
Nationality	Nominal	Categorical	Independent
Education	Ordinal	Discrete	Independent
Profession cat.	Nominal	Categorical	Independent
Work exp.	Interval	Discrete	Independent
Database ID	Interval	Discrete	Control
Employed	Nominal	Binary	Dependent

Table 4.1: Table of the chosen variables

ID	gender	birthyear	business	idea	mothertongue	finnish	english	nationality	education	profession	workexp	employed
2	male	1992	0	1	Arabic	0	4	Iraq	5	Mechanics	0	0
837	female	1953	0	0	Arabic	0	2	Iraq	4	Sciences	0	0
2897	male	1990	0	0	Kurdish	0	1	Iraq	1	Journalism	4	0
543	male	1970	0	1	Dari	2	3	Afghanista	3	Clothing	2	0
2200	female	2025	1	0	Arabic	0	0	Jordan	0	Construction	0	1
1447	male	1955	0	1	Somali	1	0	Somalia	0	Truck driver	3	0
18	female	1990	1	1		1	2	Cuba	1	Cleaning	1	0
1654	other	1993	1	1		0	3		0	Police	2	0

Figure 4.2: A sample of the possible final dataset

4.3.3 Statistical analysis

The role of statistical analysis in this study is to provide preliminary understanding of the refugees before implementing the logistic regression models.

This is done by calculating few basic distributions and key figures of the whole population and later of the employed population. Some hints towards the possible results for the research questions might already emerge here, this is why some comparisons are made between the distributions of the two populations to see whether some drastic differences exist between the two.

Distributions for all independent variables will be calculated, and for the ones with high amount of states, visualisations will be created for better interpretation in addition to numeric information (Evans and Rosenthal, 2004). Similar distributions will be calculated for both populations. Also averages for both populations will be calculated for most data of ordinal and interval scales of measurement: Finnish, English, Education level, and work experience (Evans and Rosenthal, 2004; Swatzell, K. E. & Jennnigs, 2007). In addition, a distribution for the jobs that people were employed to will also be calculated and visualized in a similar manner as the profession distributions. The visualisations will be created with WPS Spreadsheets and WPS Presentation softwares.

4.3.4 Logistic regression

A binary logistic regression is used to model the probability of an event when two possible end states exist, such as win/lose, live/die, or in this study: get employed/remain unemployed (Hosmer and Lemeshow, 2000; Peng et al., 2002). Logistic regression is similar to linear regression, but the assumption of a continuous outcome in linear regression does not apply when the outcome is binary, therefore logistic regression is the method used in this study, instead of linear regression (Peng et al., 2002). In this study logistic regression is used to model the odds of a refugee being employed as a linear combination of one or more predictor variables (Hosmer and Lemeshow, 2000). As is the case in linear regression, the predictor variables can be either binary, multicategory, discrete, or continuous (Hosmer and Lemeshow, 2000; Peng et al., 2002). Therefore logistic regression is applicable to the final dataset formed in the previous subchapter 4.3.2.

Logistic regression derives its name from the unit of measurement being modeled with a linear combination: logit, which is defined as

$$logit(p) = \ln \frac{p}{1-p}$$

, with $p = P(Y = 1)$ being the probability of the dependent variable y being 1 (Peng et al., 2002). With this in mind, we can define the logistic regression

model used in this study. Let us first define the independent variables as x_i, $i \in [1, 12]$, in this case the initial logistic regression model in this study is

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_{12} x_{12}$$

with β_i representing the regression coefficients of the model (Hosmer and Lemeshow, 2000; Peng et al., 2002).

To estimate the suitable values for β_i one must apply logistic regression with suitable data (Hosmer and Lemeshow, 2000). Unlike in linear regression, where it is possible to apply ordinary least squares to get an analytical solution to the maximum likelihood estimate (MLE) and find suitable β_i values in a straightforward manner, this does not apply to logistic regression because of its logarithmic nature and binary outcome variable (Hosmer and Lemeshow, 2000). Instead, MLE in logistic regression is solved through iterative optimisation methods instead (e.g. gradient descent or Newton's method), this in turn brings new kinds of problems (Hosmer and Lemeshow, 2000; Myung, 2003). Fitting a logistic regression model into a set of data can be problematic since the iterative process for the MLE might fail to converge (Hosmer and Lemeshow, 2000; Myung, 2003). This non-convergence might be caused by too many predictors, multicollinearity, or data sparseness among other things (Hosmer and Lemeshow, 2000). The significant case of data sparseness is something to worry with a limited dataset as the one in question, but is kept in check by cleaning and processing the data even more in the subsection 4.3.5. Concato et al. (1996) remind us that with a dataset of < 10 events per variable logistic regression's results will be unreliable and problems with convergence might emerge, so at least 10 events per variable would be preferable to avoid non-convergence.

When a statistically significant logistic regression model has been reached, the probabilities for different scenarios can be calculated from the equation describing the model itself, this technique will be applied when analysing the effect of different variables from the final model. The general equation for calculating the individual probabilities for employment naturally requires the estimation of the different β_i values, but when the estimates exist, the probabilities can be calculated with the following equation:

$$p = 1/(1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{12} x_{12})})$$

This can be further applied to analyze the effect of a single variable on employment chances by setting all but one variable x to its zero state, which is None for the profession variable, an unknown country for nationality, and a 0

for the other variables. In this case the probability function becomes simpler and calculating individual probabilities becomes more convenient:

$$p = 1/(1 + e^{-(\beta_0+\beta_i x_i)})$$

In this study the logistic regression will be implemented with the R programming language, which also provides a lot of tools for model creation and analysis, including a convenient warning for convergence errors. As mentioned before, the goal is not to create a full model for predicting refugee economical issues, but to figure out whether significant barriers exist and how severe they are for refugee employment. Because there is no specific reason to reach maximum model quality or to create a robust model for prediction, the strategy for creating the model is a stepwise logistic regression with backwards elimination, using the predictors' p-values as criteria for elimination (Hosmer and Lemeshow, 2000). In practice this means creating a model with all the predictors included, checking whether or not there are any statistically insignificant variables included, if there are: eliminating the least significant predictor and creating a new model from the remaining ones. This elimination is repeated until the model contains only statistically significant predictors or the model quality significantly plummets. Model quality during the model building is measured with the Akaike information criterion (AIC) for its convenience and ease of model comparison, lower AIC values indicate better model quality and typically when variables are eliminated, the model quality increases (Posada and Buckley, 2004).

Hosmer and Lemeshow (2000) suggest that using a 5 % significance level is too stringent for model creation and might exclude important variables, instead they suggest a 15 % significance level which is used for the model creation in this study. However, since regression models have no other way of dealing with multicategorical nominal variables but to break them down into multiple binary variables, calculating the p-values for multi category variables is not as straightforward as with continuous variables. Hosmer and Lemeshow (2000); Peng et al. (2002), and Seely and El-Bassiouni (2009) suggest using Wald test statistic for calculating the p-value for multi categorical predictors and can be considered analogous for t-tests in linear regression. On every model multicollinearity will be tested with GVIF and the variables causing possible multicollinearity will be removed, significant multicollinearity is present either if the model fails to converge or if one or more of the GVIF values exceeds 10. Goodness of fit will be measured in the final model with a Hosmer-Lemeshow test (with g = 10) to double check that the model fits the actual data well enough. The Hosmer-Lemeshow test was originally designed for logistic regression and is recommended by Hosmer and Lemeshow

(2000); Peng et al. (2002), and Bewick et al. (2005). However, Hosmer and Lemeshow (2000) leave it quite vague, how should one choose the variable g for the test, Peng et al. (2002) recommend it being at least 5, g = 10 is the deafult setting in R and it is > 5, so it is used. If the Hosmer-Lemeshow test's p-value falls under the significance level, it would indicate that the model fit is significantly inadequate (Hosmer and Lemeshow, 2000; Peng et al., 2002). After this process a model with only the significant predictors should exist. When a sensible model is created, probabilities are calculated of how severe of an effect each predictor has on refugee employment. To clarify, this logistic regression model is to model the probabilities refugees had to get employed through SUR before this research project began, which is a timespan of around 3 years. The R script used for the logistic regression modeling can be found in the A.2.

4.3.5 Data preparation for logistic regression

After the statistical analysis, the dataset still needs to be patched a little, since there are empty cells in both nationality and mother tongue columns. For the logistic regression, a distinct dataset will be used which is based on the earlier dataset with some patches and some rows deleted. Since there are only 88 rows where nationality is missing, those rows will be simply removed. After this removal, there are still 443 of those who have left their mother tongue empty. Since now every row has their nationality marked, the missing mother tongue values could be approximated by using one's nationality. However, mother tongue values can already be predicted almost perfectly from the nationality column, which implies huge multicollinearity in the upcoming model, which in turn might cause problems with convergence. Because of a lot of missing values and possible multicollinearity, the entire mother tongue column will be deleted and the amount of independent variables drops to 11. Also, because of the data sparseness the profession variable will be used in the model by its upper categories, which in this case decreases the number of possible profession categories from 56 to 11.

Even though the data has no more empty cells left, the problem that Concato et al. (1996); Hosmer and Lemeshow (2000); Peng et al. (2002); Bewick et al. (2005) all warn against: data sparseness, has not yet been taken into account. In practice this means that 1) there needs to be mostly ≥ 10 events per variable and 2) zero cells are to be avoided (Concato et al., 1996; Hosmer and Lemeshow, 2000). The first issue of general data sparseness does not seem to be a problem, since there are not too many variables in relation to

the amount of data. Zero cells are cells in a contingency table of a predictor and the dependent variable which have a zero (Pearson, 1904; Hosmer and Lemeshow, 2000). An example of an existing zero cell could be the fact that in the SUR data there is only one single person from Burundi and he has not been employed, thus in a contingency table of nationality and employment there would be a zero cell between employed = 1 and Burundi. Avoiding zero cells in this context basically means that in every category there needs to be at least 1 employed and 1 unemployed. As Hosmer and Lemeshow (2000) state about the zero cell: "This yields a point estimate for one of the odds ratios of either zero or infinity. Including such a variable in any logistic regression program causes undesirable numerical outcomes to occur.", in practice they either make the results less reliable or cause errors with convergence, both of which are to be avoided. The easiest ways to deal with zero cells are to either delete rows of data in order to remove the categories causing the zero cells or to merge entire categories together.

A contingency table is a matrix used for organising data in order to showcase the distributions of categorical variables (in this case ordinal and nominal variables) in relation to some other variable (Pearson, 1904). They are a simple way of taking a look at the basic picture between categorical variables and can also be used as a basis for calculations when only a few variables of a few possible states exist (Pearson, 1904). An example of a simple contingency table related to this study could be the distribution of men and women in the employed and unemployed populations, which can be seen in the table below. If one of the cells would be zero, it would be considered a zero cell (Concato et al., 1996).

Gender \Employed	0	1
Female	414	22
Male	2 192	193

Table 4.2: An example of a contingency table

Zero cells are typically found where not enough data exists to be distributed between the possible states of the independent and dependent variables. To locate these zero cells, before creating the logistic regression model the R script is used to create contingency tables between all categorical predictors and the dependent variable in order to spot and eliminate all possible zero cells.

Nearly all predictors had categories involving zero cells which were all

removed from the dataset: 3 rows from "other" gender, 13 Finnish native speakers, 31 rows with a doctoral degree, and 13 people with more than 8 work experiences, since no one from these categories was employed. In addition to these, the rows with nationalities with none employed or with a representation of a single row were deleted, 181 in total. After all these pieces of data were removed the total number of rows in the dataset used for logistic regression is 2 820 of which 215 were employed.

4.3.6 Predictor analysis

Even though the main logistic regression model gives us answers to the research question 1 by pointing out the most significant factors connected to refugee employment, more logistic regression models are required to answer the research question 2 - whether indirect connections to employment exist. To clarify, the results about the possible indirect connections are not as credible compared to the results about direct connections, since for indirect connections there exists significantly more alternative factors forming the path to employment than this study is able to observe. For this reason a stricter significance level of 5 % will be used, because of less amounts of available predictors and a higher level of speculation in the model. The results obtained this way are able to provide decently credible answers the research question 2.

Looking for possible indirect connections begins by building a logistic regression model from the predictors that were discarded from the main regression model of the subsection 4.3.4 as insignificant. If the second model built for the discarded predictors has any significant predictors left after the elimination rounds, it would imply a possible indirect connection with employment. To figure out the possible path of the indirect connection the predictors left after the elimination will be modeled by the predictors of the main model and each other. This should result in a situation where it is possible to perceive second order connections to refugee employment, since the connections between all variables that have a significant connection with employment have already been figured out.

To clarify the plan with an example. Let's assume that direct connections will only be perceived through education and work experience. In that case another logistic regression model would be built to model employment with all remaining predictors. Let's assume that this would show that gender has a significant indirect connection with employment. The following step would be to model gender with education and work experience as predictors, to

see whether a significant connection exists between the two. If a significant connection would exist with say education, it would imply gender having a connection with one's education and the level of education would be the factor that might affect one's probability of finding work, thus a possible indirect connection would have been spotted.

This analysis for the possible indirect connections is by no means complete to make any solid arguments of or prove the existence of a clear indirect connection. The goal is to find hints where possible indirect relationships might exist. More first order connection predictors would be required if a better analysis of the indirect relationships would be conducted. As opposed to the research of what direct connection *seem to exist* to refugee employment, this predictor analysis aims to have a look at what indirect connections to employment *might exist* through the various predictors.

4.4 Surveys

To support (or to question) the results of the logistic regression analyses, this study also includes two small surveys to gather qualitative data to be used in triangulation with the quantitative results. The purpose of these surveys is to highlight blindspots in the SUR data and check the validity of the quantitative results of this study (Mathison, 1988). The results of this qualitative part of the study will be important in compiling conclusions, since they can be used to support or question the validity of the results provided by the quantitative part (Mathison, 1988).

The aim of the two surveys is to cover the entire process of employing refugees through SUR and find factors which might block the progress within the employment process for some refugees. The first survey is targeted to the SUR employment program employees to cover the internal processes of SUR, while the second survey focuses on the employing companies' recruitment processes. Both surveys are questionnaires of only a handful of precise questions that can be answered with short answers (Owen and Noonan, 2013). The surveys were conducted through Google Forms. Both surveys will be analysed with the guidance of Thomas (2006) to first condense the answers if necessary, but more importantly to establish links between the data and the research questions of this study. The data will be grouped into different categories depending on the question, to establish the said links with the research questions and to enable comparisons between the survey results and the quantitative data analysis (Thomas, 2006; Mathison, 1988).

Since SUR's employment program employees are experienced professionals, some of them are refugees themselves, and they have detailed insider information about the entire employment process, they are the target of the first survey. The questions asked from the employees were designed to unveil the factors that would prevent a refugee from reaching a job interview: 1) Who are the people which SUR might be unable to reach or help? 2) what are the factors that help the refugees through the employment program to a job interview? This survey is sent to all staff members of the SUR employment program, 6 people in total. The questions asked in the survey are found below (with question explanations included). I take the freedom of using quite informal language in a scientific questionnaire since the persons answering the first survey are my colleagues.

1. What do you think are three kinds of people who are refugees/asylum seekers, but whom our employment program does not reach, or we are unable to help?

    - This means if you can think of people who don't want to, do not need to, or for some other reason don't use our services. If you can't think of three, then no worries, one or two is enough.

2. What do you think are the four most important qualities of a person that help them to get into a job interview through our employment program?

    - These qualities can be literally anything, from attitudes to skills, backgrounds to physical features.

The answers to the first question describe the kinds of people that do not reach SUR and therefore are probably not part of the SUR data, as well as people whom SUR might be unable to help even if they are registered to the Match system. The answers might be valuable since the descriptions can imply factors that are left outside this study, since people with some factor in a problematic state in terms of employment might not exist in the SUR data collection at all. In addition, the answers may straight out mention possible barriers for employment. The answers to the second question are more direct, since they explicitly list factors that might get people from the Match system to job interviews, therefore they might point out factors left outside this study and validate or disvalidate the factors that the logistic regression analyses will deem significant.

The answers to the first question will be categorised based on the neglected segment described by the answer. The implied barrier for employment of the segment will be noted as well. The implied barriers will then be categorised into ones supporting or questioning the results of the quantitative data analysis, and ones pointing out factors that were completely left out of this study. The implied barriers for being invited to a job interview will be listed from the answers to the question 2 and categorised as either supporting, questioning, or new factors. Similar answers to either of the questions will be combined to simplify the analysis process.

The second survey is aimed towards the employing companies to figure out their points of view for employing refugees. There are 5 companies who had recently employed people through SUR to whom the survey will be sent: Fazer, Ekovilla, N-Clean, Timanttiporaus, and Personalhuset. The companies were chosen with the help of the employment program employees. The goals of this survey are also twofold: 1) what kind of refugees (in terms of skills, profession, and quality) do the companies want to recruit? 2) What factors do the companies emphasise while recruiting? This survey contains more questions than the previous one, since I or the SUR employees do not have much information about the partner companies' recruitment processes or criteria, and the goals of the survey are more multifaceted. The questions are intentionally quite subjective and open in order to encourage personal, non-obvious viewpoints that I and SUR have not taken into account while collecting data and conducting research (Owen and Noonan, 2013).

1. What kind of employees were you looking for through Startup Refugees?
2. What were the criteria you told to Startup Refugees for finding suitable workers?
3. What are you looking for in new recruits when you conduct job interviews?
4. What separates a good recruit from a bad recruit?
5. What were the deal breakers that decide whether a person is hired or not?
6. What do you think are the most important qualities supporting or hindering employment of refugees and asylum seekers in general?

The first question is designed purely as anchoring the whole survey into the context of certain types of employees, thus there are no actual results

expected of it and it will remain outside the analysis. The questions 2 to 6 are all there to filter out various factors from different perspectives that could make a person fail a job interview, thus not getting employed, which could be included as barriers for employment. All the answers from questions 2 to 6 will be listed and categorised together. These answers will first be grouped in the more tangible/physical/quantitative and the more personality/mental/qualitative categories. However, it is highly likely that in the hugely vague category of personality related factors there will exist various kinds of factors. If easily perceivable groups emerge, they will be used to categorise the factors as well, some factors might also be applicable to multiple categories. Clearly there are infinite possible ways to do this kind of categorisation and the analysis is based on subjective pattern recognition, thus this part of the study will not provide objective results by design, which is a feature of a qualitative approach (Owen and Noonan, 2013). Finally, the barriers will be assessed whether the question, support, or imply new kinds of factors in relation to the quantitative results.

Chapter 5

Results

This chapter explains all the results of the analyses planned in chapter 4, which includes the statistical analysis, the main logistic regression analysis, the predictor analysis, and the survey results. These results are in turn used to form conclusions in the chapter 6.

5.1 Statistical analysis

This section is for exploring the overall statistics of the entire population, and making comparisons to the employed population, its purpose is to provide preliminary information for the analyses answering the research questions. The section starts by showcasing the background information of the populations, then moving into language skills, and finally work life related information.

5.1.1 Background information

The analyzed population's countries of origin were very much dominated by one single nation: Iraq, nearly representing the majority by itself. Other important countries of origin were also Afghanistan and Somalia. Below is the table 5.1 presenting the top 10 countries of origin, absolute number of SUR registered people from the country, and the relative amount of registered people. The tables include both the entire population, and the employed population for comparison. Surprisingly Iraqi people clearly dominate both populations. However, a study by Sarvimäki (2017) confirms that the refugees arriving in Europe starting from the year 2015 were dominantly Iraqis.

Whole population			Employed population		
Country	**Amount**	**Relative**	**Country**	**Amount**	**Relative**
Iraq	1494	47 %	Iraq	141	64 %
Afghanistan	479	15 %	Afghanistan	28	13 %
Somalia	232	7 %	Somalia	14	6 %
Eritrea	144	5 %	Iran	3	1 %
Syria	113	4 %	Egypt	3	1 %
Iran	77	2 %	Cameroon	3	1 %
Russia	58	2 %	Yemen	2	1 %
Turkey	29	1 %	Albania	2	1 %
Nigeria	28	1 %	Gambia	2	1 %
Yemen	26	1 %	Eritrea	1	0 %
Total	2680	85 %	**Total**	199	90 %

Table 5.1: Table of top 10 countries of origin

One can find the age distributions of the entire population and the employed population in the figures 5.1 and 5.2. What can be seen here is that the SUR clients are fairly young, a clear majority off them born in the 80's and 90's. There is a notable difference between the two populations, younger people seem to have a larger representation in the employed population compared to the whole population. The same result was also reached by Joro (2019) who also says it to be balancing the Finnish age structure. Most of the refugees are in an optimal age regarding their work life and their expected value for the society, which Hangartner and Sarvimäki (2016) say could turn out as a net positive, assuming good conditions for integration. However, SUR is basically not able to help children or youngsters since its main offering is a possible job, this means that there are basically no child profiles in Match. Therefore, the real average age of the entire refugee population is probably clearly lower, considering also that UNHCR (2019) states that 30 % of refugees arriving during 2017 were children, which is a large segment of people that are not present in SUR data.

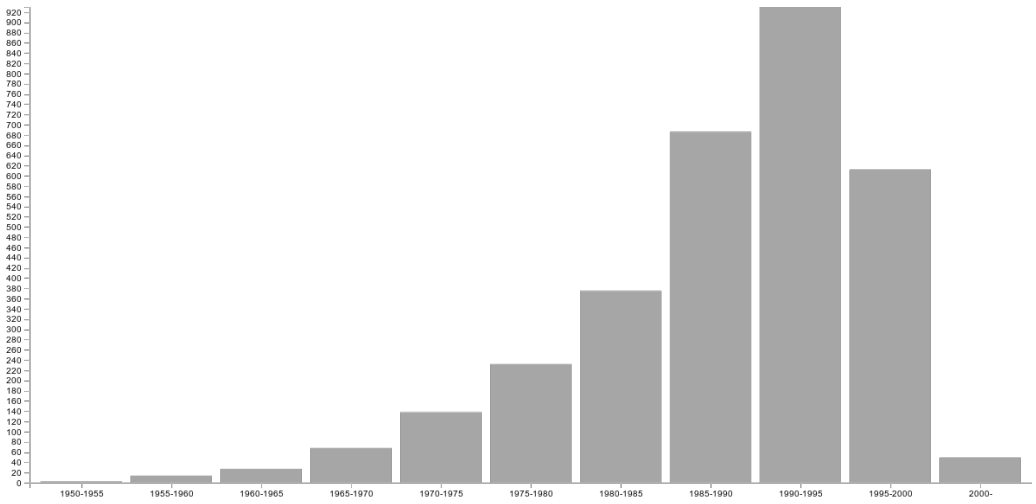

Figure 5.1: Age distribution of the whole population

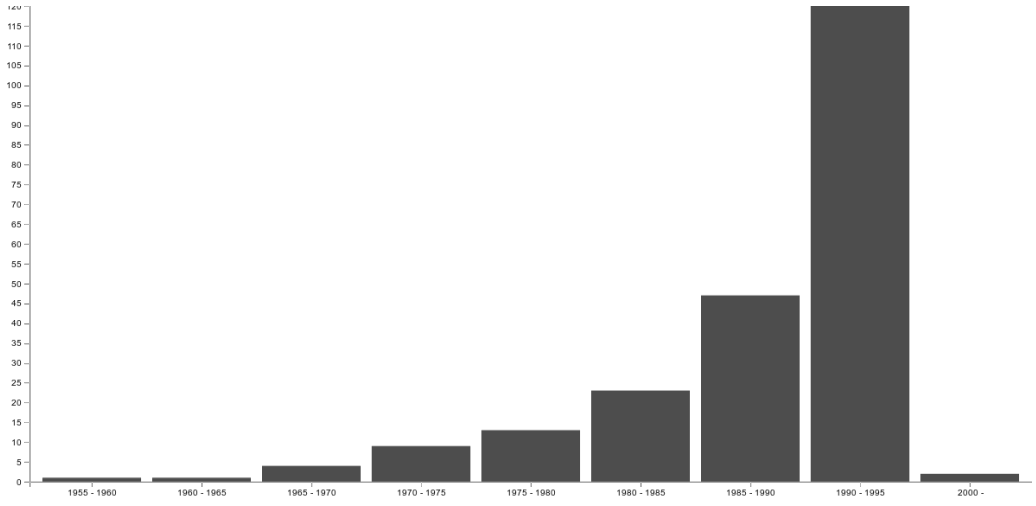

Figure 5.2: Age distribution of the employed population

Of the entire population around 83 % registered as men and around 17 % as women. 3 persons registered being of other gender. What is interesting is the difference between the two populations, women seem to be clearly underrepresented in the employed population, see table 5.2. It would have been natural to assume genders being represented evenly, that the refugee

population would be composed of a nearly 50-50 distribution of men and women, however this is clearly not the case since the distribution is roughly 95-15 where men are the majority. It could have been possible that this would be a huge distortion caused by SUR processes significantly favouring men, but the refugee overview by UNHCR (2019) backs this result by stating that of the refugees arriving in Finland during 2017 were composed of only 19 % of women.

Gender	**Whole population**	**Employed**
Male	83 %	90 %
Female	17 %	10 %

Table 5.2: Table of gender distributions

Education wise it is easy to perceive interesting differences in the two populations. One can see from the table 5.3 that the employed population has a higher education on average, but all the education level 6 persons (people with doctoral degrees) are absent from the employed population.

Education level	**Whole population**	**Employed population**
0	10 %	5 %
1	6 %	3 %
2	19 %	19 %
3	24 %	27 %
4	15 %	18 %
5	25 %	28 %
6	1 %	0 %
Average	3,0	3,4

Table 5.3: Table of education level distributions

5.1.2 Language skills

The mother tongue distribution has a lot of similarities with the nationality distribution. Arabic, Dari, and Somali form the clear majority here, in the same order and with similar representation as the top 3 nationalities: Iraq, Afghanistan, and Somalia. A similar situation as with the nationalities, some differences in the populations are perceivable, but nothing dramatic. See table 5.4 for the top 10 mother tongues of the two populations.

Whole population			Employed population		
Language	**Amount**	**Relative**	**Language**	**Amount**	**Relative**
Arabic	1397	44 %	Arabic	122	55 %
Dari	328	10 %	Dari	20	9 %
Somali	229	7 %	Somali	14	6 %
Kurdish	122	4 %	Farsi	5	2 %
Tigrinya	105	3 %	Kurdish	4	2 %
Farsi	97	3 %	Spanish	3	1 %
Russian	56	2 %	English	3	1 %
Spanish	39	1 %	Pashto	3	1 %
Pashto	33	1 %	Turkish	2	1 %
Turkish	27	1 %	French	2	1 %
Total	2450	78%	**Total**	178	81 %

Table 5.4: Table of mother tongue distributions

The average Finnish and English language skills are both higher in the employed population, a significant change can be seen in English, this would hint that English could be a possible factor with a significant connection to employment. Otherwise the English skills of the whole population are decent, Finnish is obviously a less known language. See table 5.5 for English and table 5.6 for Finnish details.

Level	**Whole pop.**	**Employed pop.**
0 - None	18 %	5 %
1 - Basic	25 %	18 %
2 - Mediocre	18 %	17 %
3 - Good	18 %	28 %
4 - Very good	20 %	31 %
5 - Mother tongue	1 %	1 %
Average	2,0	2,7

Table 5.5: Table of English language skills

Level	Whole pop.	Employed pop.
0 - None	28 %	13 %
1 - Basic	43 %	48 %
2 - Mediocre	15 %	21 %
3 - Good	11 %	12 %
4 - Very good	3 %	6 %
5 - Mother tongue	>0 %	0 %
Average	1,2	1,5

Table 5.6: Table of Finnish language skills

5.1.3 Work life information

Work experience also seems to be a possible significant factor on employment, since a notable difference exists between the two populations, as can be seen in the table 5.7, the employed population has more work experience on average.

Amount	Whole pop.	Employed pop.
0	28 %	10 %
1	27 %	18 %
2	18 %	18 %
3	13 %	29 %
4	7 %	12 %
5	3 %	8 %
6	2 %	3 %
7	1 %	1 %
>7	1 %	1 %
Average	1,7	2,6

Table 5.7: Table of work experience distributions

There is no drastic difference between the populations when observing the willingness to establish or whether one has a business idea, both statistics have a slight decrease in the employed population compared to the whole population, see table 5.8 for details.

Status	Whole pop.	Employed pop.
Wants business	39 %	37 %
Business idea	23 %	20 %

Table 5.8: Table of entrepreneurial tendencies

For clarity of this report, the information about the populations' professions will be presented in a visual format, and the numeric results will be presented only for the top 10 professions. The complete table of the professions and their numeric presentation is found in the A.2. The visual presentation of the figure 5.3 shows the distribution of the upper categories in the bigger circle, and showcases the profession distributions of the top 3 biggest upper categories in the three smaller circles. What one can see from 5.3 and 5.9 is that the construction worker profession has the most presentation, otherwise the professions are quite evenly distributed and the refugee population has people from all walks of life. There was an unexpected bias towards construction work when assigning the profession categories - the people who did not have any reported work experience or education reported themselves as wanting to do construction work, which at least partially explains the huge amount of construction workers. Unfortunately also a notably sized group without any profession exists.

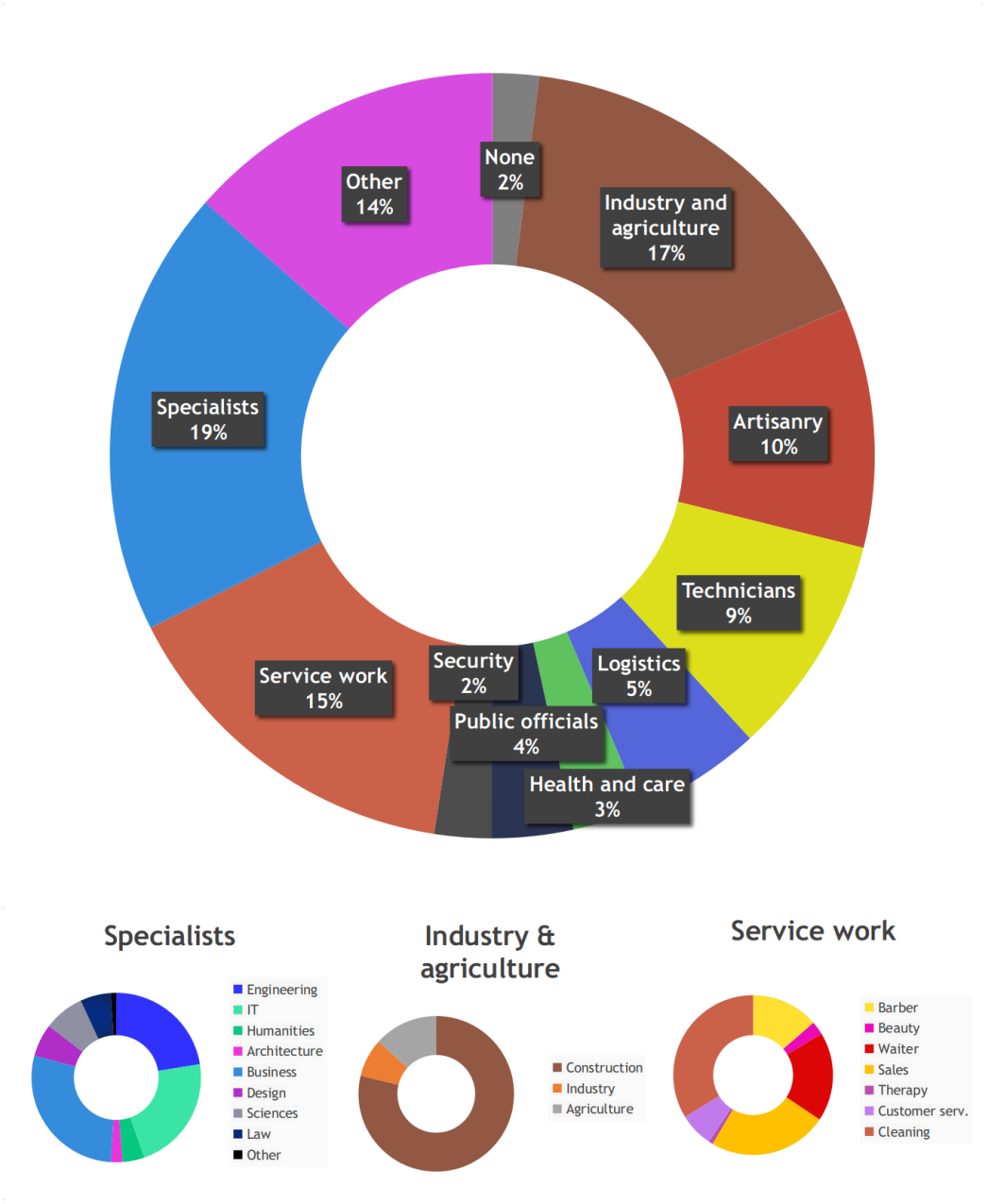

Figure 5.3: Profession distribution of the whole population

Profession	**Amount**	**Relative**
Construction	415	13,2 %
Business	168	5,3 %
Cleaning	161	5,1 %
Mechanic	157	5,0 %
Food making	144	4,6 %
Student	144	4,6 %
Engineering	134	4,3 %
IT	132	4,2 %
Sales	115	3,7 %
Clothing	92	2,9 %
Total	1662	52,9 %

Table 5.9: Table of top 10 professions

When taking a look at the employed population's professions, a significant difference can be seen: specialists, service workers, and industrial workers form a near majority by themselves, see figure 5.4 and table 5.10.

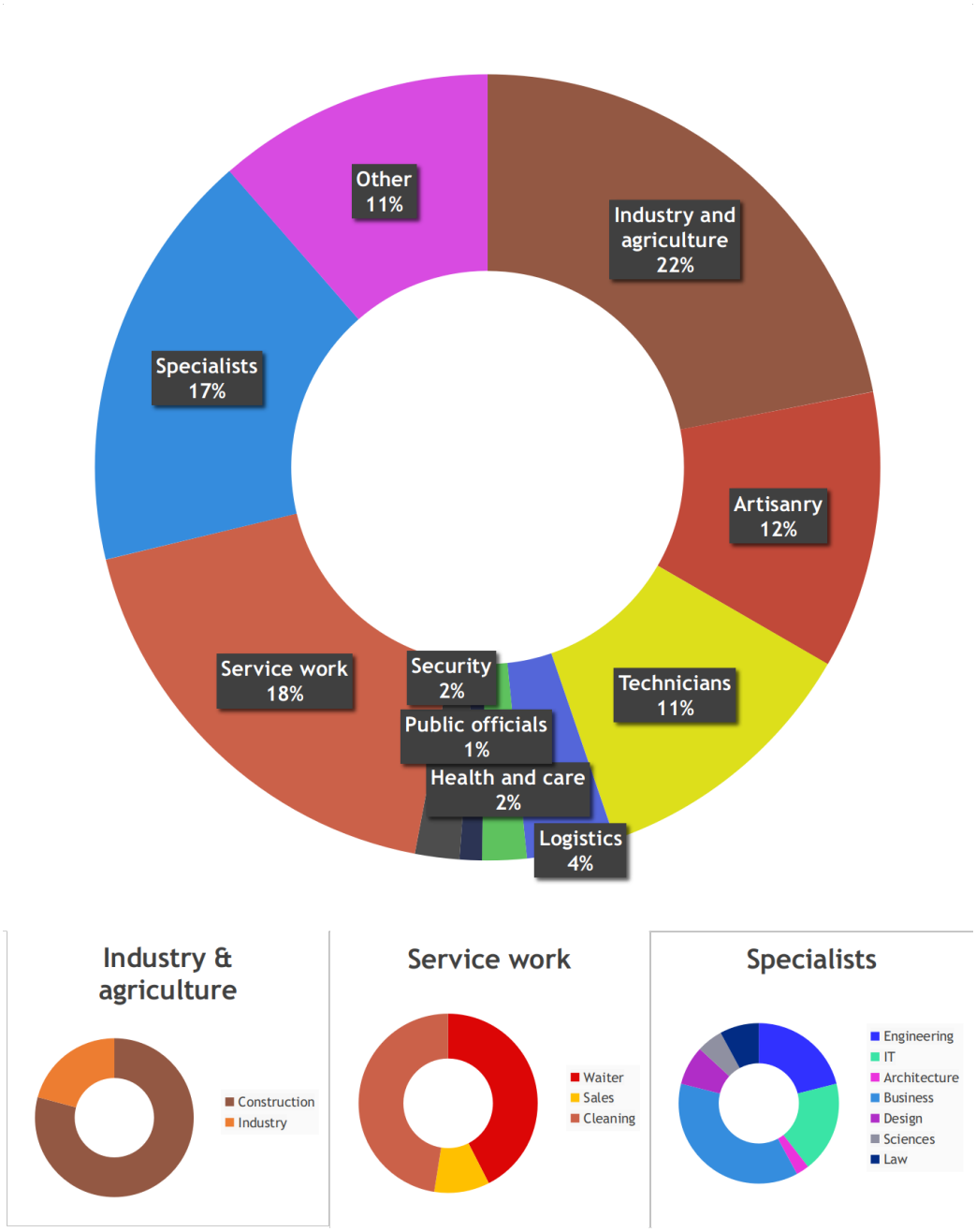

Figure 5.4: Profession distribution of the employed population

Profession	**Amount**	**Relative**
Construction	38	17,3 %
Food making	20	9,1 %
Cleaning	19	8,6 %
Waiter	17	7,7 %
Electrician	15	6,8 %
Business	14	6,4 %
Industry	10	4,5 %
Engineering	8	3,6 %
IT	7	3,2 %
Manager	6	2,7 %
Total	154	69,9 %

Table 5.10: Table of top 10 professions of the employed population

For comparison, the figure 5.5 and table 5.11 show the distribution of the jobs where the employed people has actually been employed to through SUR. This information is not important in regards to the research questions, but brings to light interesting information about SUR's performance and explains some of the findings of the profession distributions. As we can see, the employment is significantly biased towards service work and industrial labour.

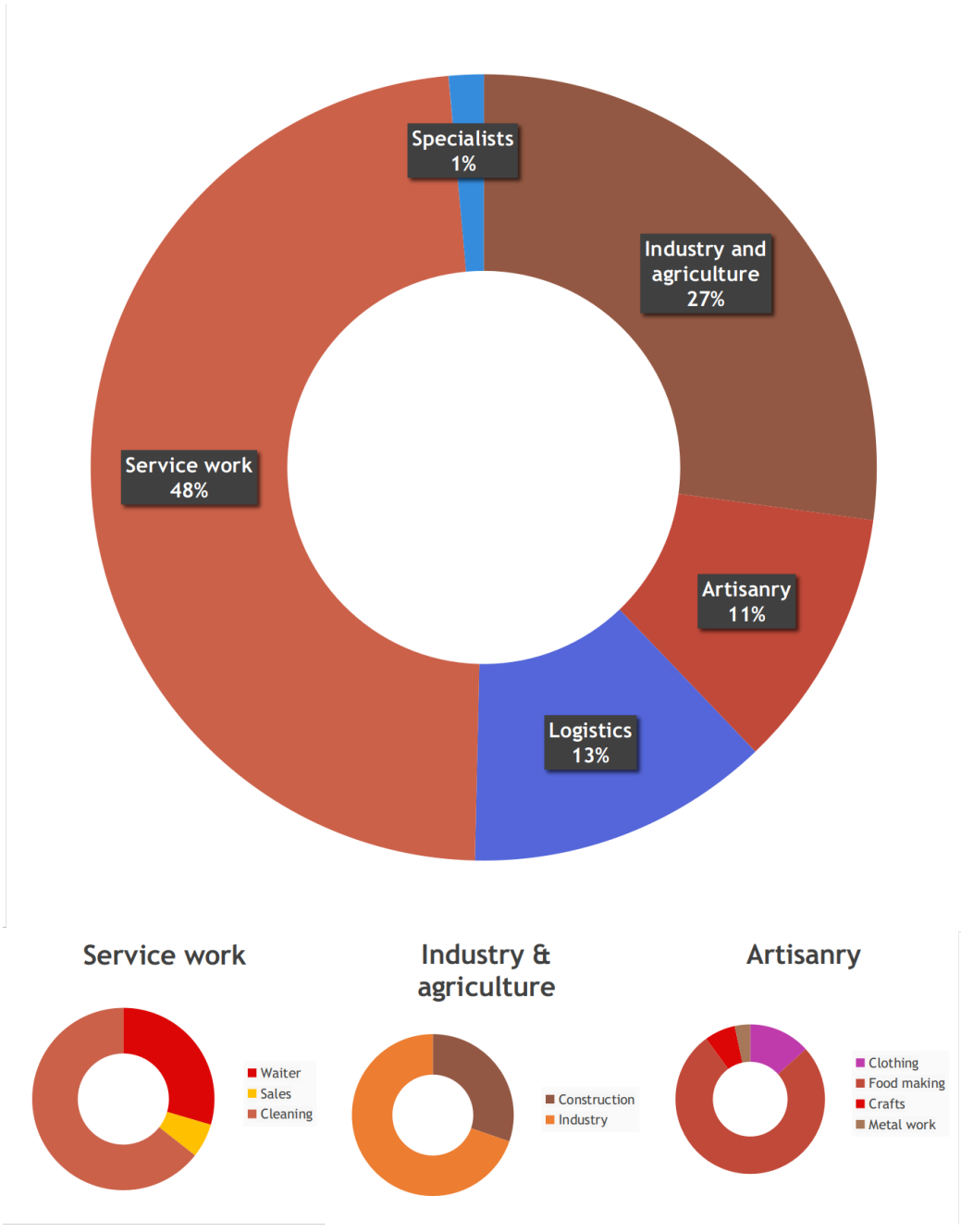

Figure 5.5: Profession distribution of the employed population

Profession	**Relative amount**
Cleaning	30,4 %
Industry	18,9 %
Waiter	13,9 %
Logistic labour	12,5 %
Construction	8,2 %
Food making	8,2 %
Sales	2,9 %
Customer service	0,7 %
Metal work	0,4 %
Therapy	0,4 %
IT	0,4 %
Total	100 %

Table 5.11: Distribution of top 10 jobs within the employed population

5.2 Logistic regression analysis

This section provides the diagnostic results which test the hypothesis made to the research question 1. The walkthrough of the model building can be found in the A.4. The fifth model of the building process was the first one where all predictors were statistically significant, thus it will also be the final model which provides answers to the research question 1. Multicollinearity was not an issue, since all GVIF values are still < 10, Hosmer-Lemeshow test implies that the model fit is on adequate levels, since its $p = 0,7607 > 0,15$, see table 5.12 for details of the model. The table showcases all regression coefficients and p-values of the chosen variables of the final model. These numeric results indicate that this model is sufficiently good and statistically significant, fit for analysis.

Predictor	β_i	**p-value**	**GVIF**
Intercept	-5,233	6,46 e-7*	
Database-ID	-1,582e-4	0,0117*	1,188
Idea	-0,3662	0,0497*	1,061
Finnish	0,1564	0,0244*	1,115
English	0,3610	1,20 e-8*	1,321
Education	0,1514	0,0206*	1,611
Profession	Various	7 e-4*	1,563
Work exp.	0,2956	1,60 e-12*	1,115
Hos.Lem.	4,971	0,7607	
AIC	1399,2		

Table 5.12: Fifth model's results. Statistical significance marked by *.

The fact that the final model's all variables have a statistically significant connection to employment implies that all of them (or their lack of) can be considered significant barriers to refugee employment in their certain states. It can thus be said with confidence that having a business idea, Finnish language skills, English language skills, education level, profession, and work experience have a significant connection to refugee employment. As a side note, this result does not mean that there cannot be any other factors, it only confirms that from the original set of factors these have a significant connection to employment, according to the SUR data. However, it would still be important to compare the effects, since not all factors are created equal and some factors' effects on employment probably are more significant than others, and some factors' effects can be negative. Also, it is interesting that the control variable database-ID was also found to be statistically significant, this implies that a model without it might not have been as accurate.

To explore the effects of the different states of each factor, the following tables showcase the sensitivity analysis of how the probabilities to get employed change compared to a certain reference state of the variable, while all other variables remain unchanged. In each table the "Probability" column describes the probability to be employed for different states of the variable, when all other variables are in their zero state. The "Relative change" column describes the relative increase or decrease of employment probability compared to the variable's zero state. The tables are shown for each variable: business idea: 5.13, Finnish 5.14, English 5.15, education 5.16, profession5.17, and work experience 5.18.

Business idea	Probability	Relative change
No	0,531 %	0 %
Yes	0,369 %	-30,5 %

Table 5.13: Connection of business idea on employment probability

Finnish	Probability	Relative change
0	0,531 %	0 %
1	0,620 %	16,8 %
2	0,724 %	36,5 %
3	0,846 %	59,4 %
4	0,988 %	86,1 %
5	1,15 %	117 %

Table 5.14: Connection of Finnish language skills on employment probability

English	Probability	Relative change
0	0,531 %	0 %
1	0,760 %	43,1 %
2	1,09 %	105 %
3	1,55 %	192 %
4	2,21 %	317 %
5	3,14 %	492 %

Table 5.15: Connection of English language skills on employment probability

Education	Probability	Relative change
0	0,531 %	0 %
1	0,617 %	16,2 %
2	0,717 %	35,1 %
3	0,833 %	57,0 %
4	0,968 %	82,4 %
5	1,12 %	112 %

Table 5.16: Connection of education level on employment probability

Profession	Probability	Relative change
None	0,531 %	0 %
Industrial & Agr.	2,28 %	329 %
Artisan	2,33 %	338 %
Technician	1,88 %	254 %
Logistic	1,04 %	96,8 %
Health & care	0,871 %	64,1 %
Public official	0,276 %	-48,0 %
Security	1,30 %	144 %
Service work	2,13 %	302 %
Specialist	0,828 %	56,0 %
Other	1,01 %	89,4 %

Table 5.17: Connection of profession on employment probability

Work exp.	Probability	Relative change
0	0,531 %	0 %
1	1,01 %	89,4 %
2	1,90 %	257 %
3	3,55 %	569 %
4	6,55 %	1 130 %
5	11,8 %	2 120 %
6	20,2 %	3 710 %
7	32,6 %	6 040 %
8	47,9 %	8 930 %

Table 5.18: Connection of work experience on employment probability

To interpret the results of the final model - it is true that all these variables seem to be statistically significant, but the effect of some variables is humongous compared to others. First of all, according to the model and the SUR data, the base probability for employment, when one has no skills in Finnish or English, no work experience, no education, no profession, and no business idea is around 0,53 %. It is easy to use this base probability as a reference point to check how big the effects of different factors are. For example, being a security worker increases one's probability to get employed from 0,53 % to 1,3 %, which is a relative increase of 144 % (table 5.17, Security row).

The dramatic effect of work experience is certainly intersting. For example: compared to Enlish language level 4 or being a professional service worker,

which have a relative probability increases of 317 % and 302% respectively and are huge increases, having 4 previous work experiences increases one's relative employment chances 1 130 %, an increase of eleven times. So one can see that even though all of the factors as variables in the final model are statistically significant, some of them have fairly minor effects, while some have a dramatically strong effects on employment chances.

There are still two things that I personally find quite interesting: 1) gender is not part of the final model while still the gender distribution difference between the two populations was quite different as seen in the table 5.2, this implies a possible indirect connection, 2) for some reason having a business idea is a significant factor and it has an effect of relative decrease of 30 % on employment probability, which can be considered weird.

5.3 Predictor analysis

The first step of the predictor analysis was to model employment with the variables left out of the main model, in this case these are: gender, birth year, willingness to establish a business, and nationality, as well as the control variable database-ID for more valid results. The walkthrough of the model building can be found from the A.5. The final model turned out to be one with only Gender as a statistically significant variable, see table 5.19. This further implies a possible indirect connection to employment, possibly through one of the variables in the main model, which will be explored next.

Predictor	β_i	**p-value**
Intercept	-2,937	<2 e-16 *
Gender male	0,5083	0,028*
Hos.Lem.	6,11 e-17	1
AIC	1518,5	

Table 5.19: Results of the fifth and final predictor model

The next step will be to model gender with all the significant predictors of the main model, the building did not actually require any iterations, since all of the variables proved statistically significant. The details of the model can be found in the table 5.20. What these results imply is that women tend to have on average less business ideas, better Finnish language, worse English language, better education, some professions, and less work experience. We

can also see from the β_i of the Work exp. and Business idea rows that the biggest differences can probably be found in the amount of work experience and having business ideas, because of its biggest β coefficient. As can be expected, the Hosmer-Lemeshow test shows by its p-value, this model's fit is not the best possible, but still good enough.

Predictor	β_i	**p-value**	**GVIF**
Intercept	0,8486	0,00653*	
Business idea	0,2746	0,0480*	1,034
Finnish	-0,1409	0,00842*	1,126
English	0,1341	0,00267*	1,367
Education	-0,1062	0,01708*	1,649
Profession	Various	∼0*	1,577
Work exp.	0,1997	5,07 e-7*	1,010
Hos.Lem.	11,21	0,19	
AIC	2109		

Table 5.20: Results of the first gender model from the male point of view

Similar to the main model's analysis, exploring the effects of the different states of each factor proceeds by examining the probabilities of the person being a man with certain reference states of different variables, while keeping all other variables remain unchanged in their zero state. In each table the "Probability" column describes the probability of the person being a man for different states of the variable, when all other variables are in their zero state. The "Relative change" column describes the relative increase or decrease of probability of being a man compared to the variable's zero state. The tables are shown for each variable: business idea: 5.21, Finnish 5.22, English 5.23, education 5.24, work experience 5.25, and profession 5.26.

Idea	**Probability**	**Relative change**
No	70,0 %	0
Yes	75,5 %	7,75 %

Table 5.21: Connection of a business idea to the probability of being male

Finnish	Probability	Relative change
0	70,0 %	0
1	67,0 %	-4,34 %
2	63,8 %	-8,89 %
3	60,5 %	-13,6 %
4	57,1 %	-18,5 %
5	53,6 %	-23,5 %

Table 5.22: Connection of Finnish language level to the probability of being male

English	Probability	Relative change
0	70,0 %	0
1	72,8 %	3,91 %
2	75,3 %	7,59 %
3	77,7 %	11,0 %
4	80,0 %	14,2 %
5	82,0 %	17,2 %

Table 5.23: Connection of English language level to the probability of being male

Education	Probability	Relative change
0	70,0 %	0
1	67,8 %	-3,25 %
2	65,4 %	-6,62 %
3	62,9 %	-10,1 %
4	60,4 %	-13,7 %
5	57,9 %	-17,4 %

Table 5.24: Connection of education to the probability of being male

Work exp.	Probability	Relative change
0	70,0 %	0
1	74,0 %	5,74 %
2	77,7 %	11,0 %
3	81,0 %	15,6 %
4	83,9 %	19,7 %
5	86,4 %	23,3 %
6	88,6 %	26,5 %
7	90,4 %	29,1 %
8	92,0 %	31,4 %

Table 5.25: Connection of work experience to the probability of being male

Profession	Probability	Relative change
None	70,0 %	0
Industry & agr.	96,4 %	37,7 %
Artisan	78,3 %	11,8 %
Technician	98,3 %	40,4 %
Logistic	99,3 %	41,8 %
Health & care	48,2 %	-31,2 %
Public official	65,3 %	-6,71 %
Security	95,2 %	35,9 %
Service work	72,0 %	2,76 %
Specialist	81,3 %	16,0 %
Other	76,2 %	8,75 %

Table 5.26: Connection of profession to the probability of being male

The details of these tables give us insight to the importance of certain variables. Most importantly, even though all of the variables of the predictor model had significant connection to one's gender, we can see that the strongest connection can be found in the amount of work experience and certain professions, namely industrial workers, technicians, lagostic workers, and security personnel (tables 5.25 and 5.26), which clearly have an increased probability of the person in question being a man compared to other factors.

5.4 Survey results

Of the two surveys the one targeted for SUR employment program's employees was a success, all 6 employees responded to the survey and the answers were thoughtful. The other survey targeted to the companies that have employed people through SUR performed only moderately, 3 out of 5 HR managers/recruiters responded and the answers were slightly more superficial than those of the SUR employee survey. This difference can be largely explained by the idea that SUR employees probably have higher motivation to answer this kind of surveys, since the survey results examines their daily work and its outcomes, also the survey comes from a colleague so there exists a more personal element to respond as well, the HR managers probably lack both of these motivations.

5.4.1 Internal survey

In this subsection the results of the internal survey will be explored by listing all the answers for both questions and categorizing them in the way described in the section 4.4. The answers will be slightly corrected, since some of them include typos or grammar flaws, and some will be slightly shortened, since some answers exist as unnecessarily long sentences. The analysed results of these surveys will be hugely important when drawing conclusions from the quantitative results, since there might be contradictions or other relationships between the two types of results that alter the validity of the results and the nature of the conclusions.

Question 1: What do you think are three kinds of people who are refugees or asylum seekers, but whom our employment program does not reach, or we are unable to help? Answers can be seen in the table 5.27.

Segment	**Implied barrier**	**Relation to results**
Illiterate people	Lack of education	Supporting
Elders	Old age	Missing Questioning
People with difficulties with transportation	Geographical distance	A new factor
Welfare trap victims	Welfare traps	Missing A new factor
Housewives	Gender + family roles	Missing Supporting
People lacking social connections	No social connections	Missing A new factor
Non-English speakers	No English language	Supporting
Non-Arabic speakers	No Arabic language	A new factor
Non-Dari speakers	No Dari language	A new factor
Highly educated	Overt education	Missing Questioning Supporting statistic
People with strong/wide social connections	Overt connections	Missing
Highly educated refugees with foreign education	Mismatch of education and job requirements	A new factor
Women	Gender	Supporting
Youngsters	Young age	Missing

Table 5.27: Categorized answers to the first question for SUR employment program employees

The first question's answers alone provide an ample amount of hints toward missing factors and data. According to these answers the connections of education, English language, and gender to refugee employment are valid results. The answers also imply that geographical distance from the workplace, Arabic, Dari, social connections, and the alignment of education and the requirements of the job might be factors that have not been taken into account by this study. Additionally the answers imply that old people, people in welfare traps, housewives/single mothers, some highly educated people, people with good social connections, and youngsters might probably be partially or completely missing from the data. This implies that some of the demographics might be biased and the logistic regression models' results might not tell

the whole truth since some people are missing from the data. The anwers might also imply that high education or old age might be possible barriers for refugee employment, which is in direct contradiction with the quantitative results which argue the insignificance of the said factors.

Question 2: What do you think are the four most important qualities of a person that help them to get into a job interview through our employment program? Answers in the table 5.28.

Category	**Implied barrier**
Supporting	Finnish language
	English language
	Work experience
	Education
Questioning	Age
New ideas	Motivation/willingness
	Being open for training periods
	Certificates
	Driving license
	Friendliness
	Hard working
	Attitude
	Behaviour
	Work life skills
	Self presentation
	Physical fitness

Table 5.28: Categorized answers to the second question for SUR internals

The question 2 answers provided more direct answers to possible barriers for employment from the point of view of being invited to a job interview. The answers supported the validity of Finnish, English, work experience, and education as significant factors to employment. However, similar to the question 1 answers, age is being highlighted as a possible barrier for employment, however the answer does not imply whether it is about young or old age or whether it depends on the job at hand. Various possible new factors also emerged, including: motivation, openness for training periods, certificates (hygiene pass, safety card etc.), driving license, friendliness, work attitude, behaviour, skills, self preservation, and physical fitness. All of these new factors could be possible factors affecting the employment chances of a refugee. Certainly interesting is the high amount of more qualitative and personal fac-

tors, related to attitude, these possible factors imply that more answers to the same research question could be found through a quantitative study about barriers for refugee employment, answers that a quantitative study like this is unable to provide.

5.4.2 Company survey

This subsection explores the results of the survey sent to the partner companies' human resource managers/recruiters. These questions in contrast to those of the other survey are much more specific and no straightforward answers or implications will be provided by the answers alone. The answers will be categorised in order to the policies described in the section 4.4 and the categorised answers can all be found in the table 5.29.

This survey managed to pinpoint a huge list of new factors, some of them related to physical features, but the clear majority of them belonging to the realm of work related psychologycal factors. As of tangible factors, different kinds of work life skills, physical fitness, and certificates were mentioned as new kinds of factors. Some work related skills and certificates were included in the SUR data, but were discarded because of the quality of data, the survey answers however imply that these might have been significant factors on employment. These answers also supported the results that work experience, English language, and Finnish language would be significant factors. No factors questioning the quantitative results were given. However, a myriad of new factors of the personality/mental/qualitative kind were perceived. Few categories which these mental factors represented were perceived: one's relationship to work, attitude, personality, mental capabilities, and social skills. Each of these categories could be interpreted as factors affecting one's chances for employment and are all completely outside the scope of this study, implying that there are numerous mental factors that might have an even more important effect on one's employment than any of the factors pinpointed by the quantitative analysis of this study.

Answer	Implied barrier	Type	Relation
Tool skills	Work life skills	Tangible	A new factor
Physical fitness	Physical fitness	Tangible	A new factor
Activity	Attitude	Attitude	A new factor
Motivation	Work motivation	Relation to work Attitude Personality	A new factor
Skills	Work life skills	Tangible	A new factor
Work experience	Work experience	Tangible	Supporting
Capability to learn	Learning	Mental capab.	A new factor
Enthusiasm	Attitude Work motivation	Attitude Relation to work	A new factor
English	English	Tangible	Supporting
Finnish	Finnish	Tangible	Supporting
Impression	Self presentation	Social skills	A new factor
Interest towards the job	Work motivation Self presentation	Relation to work Social skills	A new factor
Special skills	Work life skills	Tangible	Vague
Willingness to work after the test period	Work motivation	Relation to work	A new factor
Willingness to develop oneself at work	Work motivation Learning Attitude	Relation to work Mental capab. Attitude	A new factor
Serving the company's interest	Loyalty	Relation to work	A new factor
Satisfying customer needs	Loyalty Work life skills	Relation to work Tangible	A new factor
Rat personality type	Personality	Personality	A new factor
Lazy behaviour	Behaviour	Personality Attitude	A new factor
Trustworthiness	Trustworthiness	Relation to work Personality Attitude	A new factor
Attitude	Attitude	Attitude	A new factor
Lack of respect	Respect Loyalty	Relation to work Personality Attitude	A new factor
Work permits	Certificates	Tangible	A new factor
Responsibility	Responsibility	Personality Attitude Relation to work	A new factor
Initiative	Proactivity	Personality Attitude Relation to work	A new factor

Table 5.29: The categorised answers from questions 2 to 6

Chapter 6

Conclusions

This chapter explains what the results of the study mean in context and provides answers to the research questions. First both of the research questions are answered, followed by a synthesis which aims to provide a big picture answer as the output of this study.

6.1 Research question 1

With the results of the main model, presented in section 5.2, it is possible to argue that the amount of previous work experience has clearly the strongest connection on refugee employment, thus increases in work experience dramatically improves one's chances of employment. This implies that lack of work experience is a very significant barrier for employment. Another essential factor was English language, which did not have as dramatic of an effect compared to work experience. Since English language skills had a strong positive connection to employment, it can be deduced that the lack of English language skills are a significant barrier for employment as well, even more significant than the lack of Finnish language skill. Thus we can conclude that the hypothesis for the research question 1 - Finnish language skills being the only factor with a direct connection to employment - can be rejected.

The less important, but still statistically significant connections to refugee employment were from Finnish language skill, education, and one's profession. All of them had a positive connection to employment, except if one's profession category happens to be a public official. Public officials were the only profession category which had a lower chance of employment than even people without a profession. The most notable of the professions with the strongest positive connection to employment were industrial labourers, arti-

sans, technicians, and service workers. Therefore it can be concluded that the lack of Finnish language, education, and having a wrong profession can be considered barriers for employment as well.

What was certainly an interesting as well as a confusing result, for some reason having a business idea lowers one's chances of employment. For me personally this was troubling, since I am unable to come up with an argument of why it would interfere with one's employment in any direct way and it is not mentioned in the literature either. My personal hypothesis on it is twofold, either: 1) A portion of the people with business ideas have chosen to become entrepreneurs, therefore employing themselves and leaving no trace in the SUR database about their employment; 2) Having a business idea actually has a direct effect on refugee employment, which this study and its author are unable to comprehend. It should be accepted that the result regarding the significance of having a business idea is a legitimate result according to the SUR data. However, it remains unclear what should be concluded from this confusing result.

6.2 Research question 2

These conclusions are based on the results of the predictor analysis of the section 5.3. Some indirect connections were identified, all of them from one's gender to employment chances. The results show that gender has a strong effect on one's chances of having a lot of work experience and being in certain professions. Since work experience was by far the factor with the strongest connection to employment, we can conclude that according to the SUR data, the difference between male and female employment is mostly (but not entirely) explained by the fact that women have on average less work experience. We can also conclude that the hypothesis for the research question 2 - only gender has indirect connections to employment - was correct.

Gender was also connected to other factors as well: males were more likely to have better English language, be technicians, industrial labourers, logistic workers, and security personnel, but were also less likely to have good Finnish skills and good education compared to women. Men were also more likely to have business ideas. It can be concluded that one's gender is significantly connected to the probabilities of whether one suffers from some barriers for employment. However, since the effect of gender on employment seems to be indirect, it is unclear whether gender itself can be classified as a barrier for employment or whether it merely affects the barriers for employment.

6.3 Synthesis

To better understand the conclusions to the research questions together, see the figure 6.1 which provides a visual synthesis of the models' results. In the visual representation green arrows indicate a positive connection and red arrows indicate a negative connection, higher weight of the arrow indicates a stronger connection. For simplicity's sake, only the most significant professions are shown in the graph. As also seen in the graph, whether having a business idea should be included in the barriers for employment or not exists in the realm of maybe.

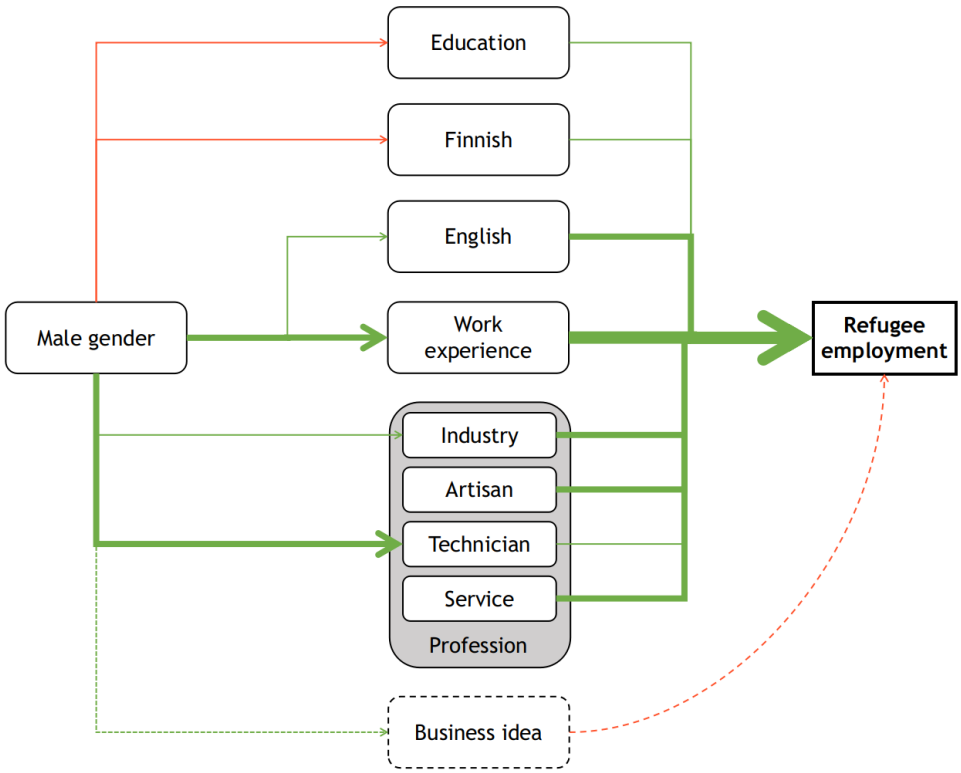

Figure 6.1: A cause and effect relationship chart of the logistic regression models' results

With the answers to the research questions in mind, it can be concluded that by having little to no work experience and poor English language skills as a refugee who does not speak Finnish, one is most certainly going to be rejected

when looking for jobs. If one happens to be female, it is more probable that one also has these characteristics for significantly harder employment.

Some contextual issues worthy of note, which cannot be directly seen from the results. In the real world refugees need to use English language with Finnish employers since most of them are unable to have conversations in Finnish. This implies that the lack of English language skills is a barrier for employment only if one cannot speak Finnish. This relationship between independent variables was left outside of this study but it might have notable effects in the real life. Also, even though lack of work experience can be considered a barrier for refugee employment, it most certainly is not a factor which makes refugees worse off compared to the native population in general. I would hypothesise that the quality of work experience is what makes more of a difference in the real world, since the problem certainly is not refugees lacking work experience, but that it is mostly from outside the Finnish job markets. Each culture has their own kinds of unique job markets, where different aspects are appreciated and emphasised. I personally assume that knowing the functioning of and habits within a certain job market is one of the most valuable kinds of work experience one could have and it is impossible to acquire it from foreign countries. Also whether the work experience is based on small family businesses or big corporations, whether it is from a culturally distant country, and whether it requires good quality basic education are aspects which I would assume that Finnish employers weigh when judging their refugee candidates. Thus the qualitative nature of the work experience might be one of the most important barriers for employment for the refugee population as a whole. One could also argue that the lack of work experience in Finland would be the real barrier for employment, however this study is unable to provide evidence for this.

The results also imply that education might not be as an important factor in refugee employment as the literature assumes after all, at least compared to language skills and work experience. However, there is the qualitative aspect in education as well, which is not taken into account in this study. It might well be, that having a non European elementary school education might be an important barrier for employment, since it is probably considered less in quality compared to the Finnish basic education. Moreover, this result implies that the lack of education probably is not the factor holding refugees in a worse off societal position, which is a common argument in political discussion. However, even though education does not seem to affect employment chances, the problem that highly educated people find it hard to get employed in their fields still remains.

Chapter 7

Discussion

This chapter considers the conclusions for this study in its grand context. First the quality of the research and the credibility of the conclusions is assessed. Next the topics needs for further academic research are examined, and finally the study's practical implications are presented.

7.1 Critique of the study

In total this study has provided information about the refugee population in Finland and concluded that certain factors contribute more or less to the chance of one's employment as a refugee. Considering the internal validity of the study, the three criteria presented by (Singleton and Straits, 2005), for a study to be able to provide sound arguments for cause and effect relationships, should be assessed. First of all, this study has provided credible proof of statistically significant connections. The connections examined in this study also have their directions of order correct by design: typically one's gender is determined before the other background factors included in this study and one first registers the background information in Match before being employed through SUR. The way the data was collected makes sure that having certain background information was the precondition for employment, not vice versa. However, this study is not by far able to eliminate rival explanations. There might be a myriad of other kinds of background factors affecting refugee employment, both qualitative and quantitative, a lot of which were mentioned in the survey results of section 5.4. Therefore more research about the topic is required to prove factors having a cause and effect relationship with refugee employment.

What one should notice about the connections between the variables.

There might be a lot more connections between the factors of this study than what was possible to perceive with the methods used. I personally believe that the factors affecting refugee employment are more like a cause and effect network rather than a few linear causal chains. This is extremely important to notice on the conclusions for research question 2, since there might be plenty of other factors through which gender might have its indirect connection to employment. This is why the results regarding the indirect connections should be taken with a grain of salt. Also, since there is probably a myriad of these direct connections to refugee employment from other factors, there are a myriad of a higher order of magnitude more factors with indirect connections to employment.

The triangulation part of this study in the section 5.4 provided a long list of possible factors that were left outside this study. Most of these factors were more of the mental variety, based on attitude, personality, and one's relationship towards work. This result implies that there is a huge amount of qualitative factors which apply to all refugees and are not as easily analysed than numerical data, but which might still have a critically important effect on employment. I would personally assume that some of them might be affected by culture, which could at least partially explain the consistently worse position of refugees all over the world. If one arrives to a new country from a distant culture, one is probably perceived different on one's attitude, personality, or relationship towards working. Most of the other results from the surveys supported the quantitative results of this study. However, the survey results pointed out that there are few refugee segments that SUR is not able to help or even reach, including but not limited to young people, old people, highly educated refugees, house wives, and single mothers. This implies that there are either some refugee segments systematically missing from the data, or some segments that are systematically struggling to get employed even if they are registered as part of the SUR employment program. Also, even the survey results should be considered incomplete, since only SUR's employees and partner companies were included. This has an effect on the sincerity of the answers, since neither target segment has an incentive to point out flaws in themselves as barriers for refugee employment. For example, the factor that employers' negative attitudes towards immigrants might be a barrier for refugee employment did not come up in the surveys even once, even though it is classified in the literature as a possible barrier.

As already mentioned in the section 4.2, this study's way of data collection is far from optimal in terms of reliability. Most of the reliability problems arise from the fact that if all of the Match profiles would be filled all over again, a

lot of the data would not be the same. Even though the variables were chosen to be as realistic as possible, even the reliability of most of the variables cannot be guaranteed. Language skills of a person can and probably will change over time, and some of the data might be based on exaggeration. There was also no way of reliably distinguish between refugees, asylum seekers, immigrants, and second generation immigrants, which also undermines the reliability of the results. Therefore, if this kind of study is conducted somewhere else with a different data collection method the results might be different. Moreover, the fact that the SUR's data about the mediated employments is incomplete is probably a biased mistake, which also bias the results. As mentioned in the section 6.1, at least there is data missing about employed entrepreneurs which might have created a biased result that having a business idea reduces chances of employment. Based on the fact that the employment dataset is incomplete, there is a possibility of more biased imperfections in the data.

Regarding the external validity of this study, the sample size of 3149 used in this study might well be big enough for the conclusions to be generalisable to the entire refugee population in Finland of around 30 000. However, the fact that this study focused on refugees' employment through SUR and not the independently alters the external validity. How SUR reaches its client refugees, who they are able to serve, and what the partner organisations are at least partially define the results of this study. Considering other environments than Finland, the factors examined in this study were very universal, thus these results could possibly be generalised to another European setting as well if one only replaces the Finnish language variable with native language. However, more research is required to further validate these kinds of results before any generalisation should be made.

7.2 Further research

Considering the rival explanations that this study was not able to eliminate, this topic requires plenty of further research to figure out more factors that might effect the employment of refugees. Studies of this kind would help to provide more evidence, to enable sound arguments for specific cause and effect relationships. However, all this requires more data about refugess which is definitely not easy to come up with. Based on my personal best knowledge, the lack of data is the main reason why evidence about this topic is hard to come by in the first place. I would personally encourage SUR among other refugee organisations and national refugee institutions to collect more data about refugees in order to study their integration and employment further. It

would also be beneficial to have results about the same topic from outside SUR context, in order to eliminate the biases caused by missing data and missing segments of people.

Once a point could be reached that sufficient digital data about refugees exists for further research, this data would be an excellent opportunity to apply machine learning algorithms. If analyses like the ones implemented in this study should be done numerous times over various factors with huge amounts of data, machine learning would make the analysis process significantly more efficient and make it possible to observe the topic more broadly. This kind of machine learning approach with sufficient amounts of data could make it practical to start mapping the topic of refugee employment more in the sense of a cause and effect network, rather than causal chains. One could also argue that this study could have also been implemented through machine learning algorithms to reach the same results, which is probably true.

Something which was already started by Joro (2019), this topic requires further qualitative studies venturing further into the attitude and personality related factors affecting employment and integration. Considering the importance that mental factors hold in the survey results, I would argue that critically important knowledge about refugee integration and employment could be uncovered by studying the mental factors related to employment.

7.3 Practical implications

The conclusions provide ground for few arguments that call for change in the current Finnish refugee politics. The result that work experience has the biggest effect of the factors in this study implies that the idea of rapid employment is more important than ever: refugees should be employed rather sooner than later. Not only because it would save tax money for other uses, but because one could have valuable work experience from the new home country, that could then be leveraged further down the refugee's career. Rapid employment in a new country would radically improve the probability of a refugee staying on his career for the rest of his life, rather than falling to unemployment at some point. However, it should be acknowledged that approaching the rapid employment of refugees questions the 3 or 6 month waiting time before refugees and asylum seekers can legally start working. Finland should consider whether this time limit is actually beneficial. It has also been found out that the chances of refugees getting employed drops the longer one needs to stay unemployed. Considering the young age of most of the refugees, this issue is of utmost importance. To not supporting the rapid employment of

refugees would be to waste tax money on unemployment support and keep talented people in their prime working age out of of the job markets. This waste of talent, work force, and people's time is hurting both the refugees and the nation.

In terms of refugee employment, English language skills seems to be significantly more important than Finnish language. This implies that the education offered to new refugees and asylum seekers should also focus on English language in addition to Finnish. English is also for most refugees the easier language to learn, so it could definitely be the easier way to faster employment. However, it would be easy to argue based on the results the education of Finnish should be entirely replaced with English studies. This argument overlooks the fact that even though English is important for employment, Finnish language might be hugely important in other facets of life. What also makes English teching important? If a refugee is forcibly or by his own decision moved to another country, English language skills are probably going to be an extremely valuable asset anyway.

The results show men having a significant edge on employment mainly because of higher average work experience and English skills. This makes it even more important to focus on women empowerment and offering job opportunities to women as well. It would also be important for the arriving refugee families to know that in Finland women and men have near equal opportunities in the job market, and that the role of women is no longer taking care of the home. Nearly all professions are open for both genders. In addition, families should be actively informed about the finnish kindergarten and education systems, so having children would not be the sole reason for staying only at home. In addition, men need this information as well in order to form shared understanding within families about the roles of both parents.

A fourth implication arises mainly from the statistical outlook of the subsection 5.1.3. One can noticed that refugees with higher education tend to have slightly higher chances for employment than people with lower education. We can also notice that in the employed refugee population the highly educated people are fairly well represented. However, what we can also see is that in the jobs where people have been employed through SUR, the highly educated jobs are almost completely missing. This implies that there are highly educated refugees being employed in Finland to jobs where their talent is going to waste. This effect has been noted by Krahn et al. (2007), in Finland by Joro (2019), and also in the survey results by the SUR's employees. The situation of the refugees is usually desperate enough that one has to take whatever job is at hand. This is also possibly affected by a mismatch

of Finnish and foreign education standards - expertise from the Iraqi education system might not be directly applicable in Finland. This finding implies that there should be courses available where educated refugees could fill in the missing pieces of knowledge to be able to join the specialist work force in Finland. It is clear that this is a problem where the help of educational institutions is required. This also implies that SUR has a fairly big segment of people which it is unable to help and is a specific topic where the SUR's model requires significant improvement.

If the Finnish government wishes to achieve its goals of successful refugee integration, I would argue based on the conclusions of this study and the established literature, that the rapid employment of refugees should be the basis of the integration process. The rapid employment should be facilitated by focusing on the refugees' language skills, supplementing the education of specialist refugees to be on par with Finnish standards, and educating refugee women about their possibilities in a new culture. This approach will not fix the worse economic status of refugees, but would make it easier for both refugees and Finland to manage when masses of people fleeing from their countries want a safe place to stay.

Refugees as a phenomenon are probably not going to disappear from the world any time soon. The conflicts of the 2015 refugee crisis are still going on, new possible conflicts can be seen in the near future, and climate refugees are already becoming a phenomenon of this century. Learning from the past events can make it easier to cope with the humanitarian crises in the future. The approach presented in this study could make it easier for both nations and refugees to get through hard times more easily. It would benefit the refugees by giving them a chance for an independent life, while freeing up societal resources for improving the lives of everyone.

This report will be used as a basis for a shorter, simpler report which is going to be published by Startup Refugees. The report will be targeted to the Finnish city and municipal administrations, interested politicians, and other refugee organisations, but will certainly be publicly available for all interested individuals and organisations.

Bibliography

Aycan, Z. and Berry, J. W. (1996). Impact of employment-related experiences on immigrants' psychological well-being and adaptation to Canada. *Canadian Journal of Behavioural Science*, 28(3):240–251.

Berry, J. (1992). Acculturation and adaptation in a new society. *International Migration*, 30.

Bevelander, P. and Lundh, C. (2007). Employment integration of refugees: The influence of local factors on refugee job opportunities in Sweden. *IZA Discussion Papers*, (2551).

Bewick, V., Cheek, L., and Ball, J. (2005). Statistics review 14: Logistic regression. *Critical Care*, 9(1):112–118.

Borjas, G. J. (1994). The Economics of Immigration. *Journal of Economic Literature*, 32:1667–1717.

Borjas, G. J. (2013). *Labor Economics*. McGraw-Hill, New York, 6 edition.

Chiswick, B. R. (1978). The Effect of Americanization on the Earnings of Foreign-born Men. *Journal of political Economy*, 86(5):897–921.

Concato, J., Feinstein, A. R., Peduzzi, P., Kemper, E., and Holford, T. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 49(12):1373–1379.

Desiderio, M. V. (2016). Integrating Refugees into Host Country Labor Markets: Challenges and Policy Options. *Washington, DC: Migration Policy Institute.*

Easterbrook, S., Singer, J., Storey, M.-A., and Damian, D. (2008). Selecting Empirical Methods for Software Engineering Research Guide to Advanced Empirical Software Engineering. *Guide to Advanced Empirical Software Engineering*, pages 285–311.

Eronen, A., Härmälä, V., Jauhiainen, S., Karikallio, H., Karinen, R., Kosunen, A., Laamanen, J.-P., and Lahtinen, M. (2014). Maahanmuuttajien työllistyminen. Technical report, Työ- ja elinkeinoministeriö, Helsinki.

European Court of Human Rights (1971). European Convention on Human Rights. *International and Comparative Law Quarterly*, 20(4):766–768.

Evans, M. J. and Rosenthal, J. S. (2004). Probability and Statistics The Science of Uncertainty Second Edition.

Feeney, A. (2000). Refugee Employment. *Local Economy*, 15(4):343–349.

Finnish Immigration Service (2019a). Alien's passport. Available online at https://migri.fi/en/alien-s-passport.

Finnish Immigration Service (2019b). Asylum seeker's right to work. Available online at https://migri.fi/en/asylum-seeker-s-right-to-work.

Finnish Immigration Service (2019c). Negative decision. Available online at https://migri.fi/en/asylum-in-finland/negative-decision.

Finnish Immigration Service (2019d). Processing Time Checker. Available online at https://migri.fi/en/check-the-processing-time-of-your-application.

Finnish Immigration Service (2019e). Residence permit application for an employed person (TTOL). Available online at https://migri.fi/en/residence-permit-for-an-employed-person.

Finnish Immigration Service (2019f). Tilastot. Available online at http://tilastot.migri.fi.

Hainmueller, J., Hangartner, D., and Lawrence, D. (2016). When lives are put on hold: Lengthy asylum processes decrease employment among refugees. *Science Advances*, 2(8).

Hangartner, D. and Sarvimäki, M. (2016). Dealing with the Refugee Crisis : Policy Lessons from Economics and Political Science. *Mimeo*.

Haverinen, V.-S. (2018). Pakotetusta toimettomuudesta - turvapaikanhakijoiden kotoutuminen vastaanottotoiminnan osana. *Janus Sosiaalipolitiikan ja sosiaalityön tutkimuksen aikakauslehti*, 26(4):309–325.

Hiekkavuo, A. (2016). Population with foreign background in Helsinki. Technical report, City of Helsinki Urban Facts, Helsinki.

Hosmer, D. W. and Lemeshow, S. (2000). *Applied Logistic Regression*. John Wiley & Sons, Inc., 2nd editio edition.

Isphording, I. E., Otten, S., and Sinning, M. (2014). The Effect of Language Deficiency on Immigrant Labor Market Outcomes in Germany. *Paper presented at the 11th SOEP User Conference*.

ja hyvinvoinnin laitos, T. (2019). Maahanmuuttajien mielenterveys. Available online at https://thl.fi/fi/web/maahanmuutto-ja-kulttuurinen-moninaisuus/maahanmuutto-ja-hyvinvointi/maahanmuuttajien-mielenterveys.

Jick, T. D. (1979). Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly*, 24(4):602.

Joro, T. (2019). Turvapaikanhakijoiden osaaminen yhteiskunnallisena voimavarana. Technical report, University of Eastern Finland, Joensuu.

Krahn, H., Derwing, T., Mulder, M., and Wilkinson, L. (2007). Educated and underemployed: Refugee integration into the Canadian labour market. *Journal of International Migration and Integration / Revue de l'integration et de la migration internationale*, 1(1):59–84.

Kurzgesagt - In a Nutshell (2017). Universal Basic Income Explained - Free Money for Everybody? UBI. Available online at https://youtu.be/kl39KHS07Xc.

Legrain, P. (2016). Refugees work: A humanitarian investment that yields economic dividends. Technical report, Tent.

Legrain, P. (2017). How to get refugees into work quickly. *Step up*, (August).

Mathison, S. (1988). Why Triangulate. *Educational Researcher*, 17(2):13–17.

Mill, J. S. (1879). *Utilitarianism*. Longmans, Green and Co., London, 7 edition.

Ministry of the Interior - Finland (2019). Refugees and asulym seekers. Available online at https://intermin.fi/en/areas-of-expertise/migration/refugees-and-asylum-seekers.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47(1):90–100.

Owen, D. and Noonan, M. (2013). Preparing and conducting interviews to collect data. Doody O, Noonan M. *Nurse researcher*.

Pearson, K. (1904). Mathematical Contributions to the Theory of Evolution. In *Drapers' Company Research Memoirs*, volume 1, chapter XIII. On t, page 43. London.

Peng, C. Y. J., Lee, K. L., and Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *Journal of Educational Research*, 96(1):3–14.

Peri, G. and Sparber, C. (2008). Task specialization, immigration, and wages. *American Economic Journal: Applied Economics*, 1(3):135–169.

Posada, D. and Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Systematic Biology*, 53(5):793–808.

Saarela, P. (2017). Euroopan komissio tukee hylkäävän päätöksen saaneiden turvapaikanhakijoiden säilöönottoa. Available online at https://www.kansalainen.fi/euroopan-komissio-tukee-hylkaavan-paatoksen-saaneiden-turvapaikanhakijoiden-pidatyksia/.

Sarvimäki, M. (2010a). Maahanmuuton taloustiede: lyhyt johdatus*. *Economist*, pages 253–270.

Sarvimäki, M. (2010b). Vievätkö maahanmuuttajat kantaväestön työpaikat? *Talous ja Yhteiskunta*, 4:10–15.

Sarvimäki, M. (2011). Assimilation to a welfare state: Labor market performance and use of social benefits by immigrants to Finland. *Scandinavian Journal of Economics*, 113(3):665–688.

Sarvimäki, M. (2017). Labor Market Integration of Refugees in Finland. Technical report, VATT Institute for Economic Research.

Schutt, R. and O'Neil, C. (2013). *Doing data science: Straight talk from the frontline*. O'Reilly Media, Inc.

Seely, J. F. and El-Bassiouni, Y. (2009). Applying Wald ' s Variance Component Test. *The Annals of Statistics*, 11(1):197–201.

Shields, M. A. and Price, S. W. (1999). The English Language Fluency and Occupational Success of Ethnic Minority Immigrant Men Living in English Metropolitan Areas. *Journal of Population Economics*, 15(1).

Siironen, S. (2019). Punainen Risti: Turvapaikanhakijoilla monipuolista osaamista, muttei pääsyä Suomen työelämään - "Mikä vain työ käy". Available online at https://yle.fi/uutiset/3-10781804.

Singleton, R. and Straits, B. (2005). *Approaches to social research*. Oxford University Press, New York, 4th edition.

Sisäasiainministeriö (2011). Laki kansainvälistä suojelua hakevan vastaanotosta sekä ihmiskaupan uhrin tunnistamisesta ja auttamisesta. Available online at https://www.finlex.fi/fi/laki/ajantasa/2011/20110746.

Smith-Spark, L. (2015). European migrant crisis: A country-by-country glance. Available online at https://edition.cnn.com/2015/09/04/europe/migrant-crisis-country-by-country/.

Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science*, 103(2684):677–680.

Swatzell, K. E. & Jennnigs, P. (2007). Descriptive research: The nuts and bolts. 20(July):3–4.

Thomas, D. R. (2006). A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation*, 27(2):237–246.

Tilastokeskus (2018). Sukupuolten tasa-arvo suomessa 2018.

Timsit, A. (2017). 'Things Could Get Very Ugly' Following Europe's Refugee Crisis. Available online at https://www.theatlantic.com/international/archive/2017/10/qa-sasha-polakow-suransky-immigration-europe/543537/.

UNHCR (1967). Convention and Protocol Relating to the Status of Refugees.

UNHCR (2019). Overview Northern Europe. Technical report, UNHCR. Available online at https://www.unhcr.org/neu/17855-statistics-refugees-asylum-seekers-northern-europe-region.html.

Valtioneuvosto (2015). Hallituksen turvapaikkapoliittinen toimenpideohjelma. Available online at https://valtioneuvosto.fi/documents/10184/1058456/Hallituksen+turvapaikkapoliittinen+toimenpideohjelma+8.12.2015/98990892-c08e-4891-8c23-0d229f1d6099.

Yijälä, A. (2014). Maahanmuuttajien taloudellinen sopeutuminen Pohjoismaissa. Maassaolon keston ja muiden aikasidonnaisten tekijöiden vaikutus tukitarpeeseen. Available online at https://www.hel.fi/hel2/Tietokeskus/julkaisut/pdf/14_11_18_Tyopapereita2_Yijala.pdf.

Yijälä, A. and Nyman, M. (2017). Living in Limbo: Qualitative case study of skilled Iraqi asylum seekers in Finland. Technical report, City of Helsinki Urban Facts.

Yle (2019). Study: 4 in 10 asylum seekers in Finland report anxiety, depression. Available online at https://yle.fi/uutiset/osasto/news/study_4_in_10_asylum_seekers_in_finland_report_anxiety_depression/10837133.

Appendix A

Appendices

A.1 Appendix 1

A list of all the professions and profession categories used in this study, with explanations included:

None: None

Industry & agriculture: Construction Regular workers, welders, and low ranking supervisors

Industry Factory and production line workers

Agriculture Farmers, foresters, fishers, cattle raisers, and herders

Artisans: Carpentery

Clothing Tailors, dressmakers, and shoemakers

Food production Cooks, chefs, brewers, butchers, and food production workers

Crafts

Metal work

Technicians: Includes professions that could overlap with construction, but typically require more specialised education

Electricity

Plumbing

Mechanic

Repairs

Logistics: Includes everything regarding moving things from A to B

Car driver

Truck driver Includes also bus drivers

Sailor Boat or ship crew

Industrial machine operator Crane drivers, mining machine operators, and operators of industrial machinery that require special training

Logistic labourer Workers of storehouses and shipyards, also delivery jobs like couriers and postmen

Health & care: Doctor

Nurse

Dentist

Elderly care

Child care Both institutionalized and individual caretakers

Pharmacy

Animal care Workers of stables', kennels, and veterinaries

Public officials: Professions that are typically available from the public sector and specialised in a particular topic

Teacher

Public official Mayors, tax officials, urban planners, common bureaucrats

Security: Soldier Regulars, special operatives and officers

Police

Guard

Rescue Lifeguards and ambulance drivers

Service work Barber Barbers, bearders, and hairdressers

Beauty Cosmetologists, pediatrics, and makeup artists

Waiter All restaurant staff that is not making food

Sales Both big business sales deals and shop clerks

Therapy Aroma, psycho, art, and ayurveda therapists, masseauses and chiropractics

Customer service Also clerks of more specialised companies like pharmacies, gyms, hotels, etc.

Cleaning Business, household, industrial, or construction cleaning

Specialists: All professions related to higher education, research, and academia

Engineering The more traditional engineering: electrical, mechanical, robotics, aeronautics, chemical, etc.

IT IT, computer science, programming, software engineering

Humanities

Architecture

Business

Design

Sciences

Law

Other Everyone else that did not fit into the other specialist professions

Other: Student

Entrepreneur

Artist

Sports

Translation

Journalism

Media Workers in media production, who are not necessary journalists or artists

Manager Supervisors, manager, coordinators, leaders and executives

Religious

A.2 Appendix 2

The R code used for this study

```
#Set the workspace ready for use
rm(list=ls());
setwd("calculations/R");
```

```
library(readxl);
library(car);
library(aod);
library(ResourceSelection);
masterdata <- read_excel("logitdata.xlsx");

#Categorize professions
for (i in 1:dim(masterdata[1])[1]){
  if (masterdata[i,10] >= 1 & masterdata[i,10] <
   10) masterdata[i,10] <- 1
  else if (masterdata[i,10] >= 10) masterdata[i,10] <-
   floor(masterdata[i,10] / 10) * 10
  else masterdata[i,10] <- 0;
}
masterdata$profession <- factor(masterdata$profession);

#Test for zero cells
# zero1 <- table(masterdata$gender, masterdata$employed);
# zero2 <- table(masterdata$business, masterdata$employed);
# zero3 <- table(masterdata$idea, masterdata$employed);
# zero4 <- table(masterdata$finnish, masterdata$employed);
# zero5 <- table(masterdata$english, masterdata$employed);
# zero6 <- table(masterdata$nationality, masterdata$employed);
# zero7 <- table(masterdata$education, masterdata$employed);
# zero8 <- table(masterdata$profession, masterdata$employed);
# zero9 <- table(masterdata$work, masterdata$employed);



#Start building models
model1 <- glm(employed ~ id + gender + birthyear + business +
 idea + finnish + english + nationality + education +
 profession + work, data = masterdata, family = "binomial");
summary(model1)
#Test nationality
wald.test(b = coef(model1), Sigma = vcov(model1), Terms =
9:33)
#Test profession
wald.test(b = coef(model1), Sigma = vcov(model1), Terms =
```

```
35:44)
vif(model1)

#Without Birthyear
model2 <- glm(employed ~ id + gender + business + idea +
finnish + english + nationality + education + profession +
work, data = masterdata, family = "binomial");
summary(model2)
#Test nationality
wald.test(b = coef(model2), Sigma = vcov(model2), Terms =
8:32)
#Test profession
wald.test(b = coef(model2), Sigma = vcov(model2), Terms =
34:43)
vif(model2)

#Without business
model3 <- glm(employed ~ id + gender + idea + finnish +
english + nationality + education + profession +
work, data = masterdata, family = "binomial");
summary(model3)
#Test nationality
wald.test(b = coef(model3), Sigma = vcov(model3), Terms =
7:31)
#Test profession
wald.test(b = coef(model3), Sigma = vcov(model3), Terms =
33:42)
vif(model3)

#Withour gender
model4 <- glm(employed ~ id + idea + finnish + english +
nationality + education + profession + work, data =
masterdata, family = "binomial");
summary(model4)
#Test nationality
wald.test(b = coef(model4), Sigma = vcov(model4), Terms =
6:30)
#Test profession
wald.test(b = coef(model4), Sigma = vcov(model4), Terms =
```

```
32:41)
vif(model4)

#Without nationality
model5 <- glm(employed ~ id + idea + finnish + english +
education + profession + work, data = masterdata, family =
 "binomial");
vif(model5)
hoslem.test(masterdata$employed, fitted(model5), g=10)
summary(model5)
#Test profession
wald.test(b = coef(model5), Sigma = vcov(model5), Terms =
7:16)


#Propabilities into arrays:
idea <- array(0, dim=c(2,2));
for (i in 1:2){
  idea[i, 1] <- 1/(1 + exp(-(coefficients(model5)[3] * (i -
   1) + coefficients(model5)[1])));
  idea[i, 2] <- idea[i, 1] / idea[1, 1] - 1;
}

finnish <- array(0, dim=c(6,2));
for (i in 1:6){
  finnish[i, 1] <- 1/(1 + exp(-(coefficients(model5)[4] *
   (i - 1) + coefficients(model5)[1])));
  finnish[i, 2] <- finnish[i, 1] / finnish[1, 1] - 1;
}

english <- array(0, dim=c(6,2));
for (i in 1:6){
  english[i, 1] <- 1/(1 + exp(-(coefficients(model5)[5] *
   (i - 1) + coefficients(model5)[1])));
  english[i, 2] <- english[i, 1] / english[1, 1] - 1;
}

education <- array(0, dim=c(6,2));
for (i in 1:6){
```

```
  education[i, 1] <- 1/(1 + exp(-(coefficients(model5)[6] *
   (i - 1) + coefficients(model5)[1])));
  education[i, 2] <- education[i, 1] / education[1, 1] - 1;
}

profession <- array(0, dim=c(11,3));
profession[1, 2] <- 1/(1 + exp(-(coefficients(model5)[1])));
profession[1, 1] <- 0;
profession[2, 1] <- 1;
profession[3, 1] <- 10;
profession[4, 1] <- 20;
profession[5, 1] <- 30;
profession[6, 1] <- 40;
profession[7, 1] <- 50;
profession[8, 1] <- 60;
profession[9, 1] <- 70;
profession[10, 1] <- 80;
profession[11, 1] <- 90;
for (i in 2:11){
  profession[i, 2] <- 1/(1 + exp(-(coefficients(model5)[5 +
   i] + coefficients(model5)[1])));
  profession[i, 3] <- profession[i, 2] / profession[1, 2] - 1;
}

work <- array(0, dim=c(9,2));
for (i in 1:9){
  work[i, 1] <- 1/(1 + exp(-(coefficients(model5)[16] *
   (i - 1) + coefficients(model5)[1])));
  work[i, 2] <- work[i, 1] / work[1, 1] - 1;
}


#Predictor analysis
predictor1 <- glm(employed ~ id + gender + birthyear +
business + nationality, data = masterdata, family =
"binomial");
vif(predictor1)
summary(predictor1)
wald.test(b = coef(predictor1), Sigma =
```

```
vcov(predictor1), Terms = 6:30)

#Without business
predictor2 <- glm(employed ~ id + gender + birthyear +
nationality, data = masterdata, family = "binomial");
vif(predictor2)
summary(predictor2)
wald.test(b = coef(predictor2), Sigma =
vcov(predictor2), Terms = 5:29)

#Without birthyear
predictor3 <- glm(employed ~ id + gender +
nationality, data = masterdata, family = "binomial");
vif(predictor3)
summary(predictor3)
wald.test(b = coef(predictor3), Sigma =
vcov(predictor3), Terms = 4:28)

#Without ID
predictor4 <- glm(employed ~ gender +
nationality, data =  masterdata, family = "binomial");
vif(predictor4)
summary(predictor4)
wald.test(b = coef(predictor4), Sigma =
vcov(predictor4), Terms = 3:27)

#Without nationality
predictor5 <- glm(employed ~ gender, data =
masterdata, family = "binomial");
summary(predictor5)
hoslem.test(masterdata$employed, fitted(predictor5), g=10)

#Other stuff against gender
genderdata <- masterdata
for (i in 1:dim(genderdata[1])[1]){
  if (genderdata[i,2] == "male") genderdata[i,2] <- 1
  else if (masterdata[i,2] == "female") genderdata[i,2] <- 0;
}
genderdata$gender <- as.numeric(genderdata$gender);
```

```
gend1 <- glm(gender ~ idea + finnish + english + education +
profession + work, data = genderdata, family = "binomial");
summary(gend1)
wald.test(b = coef(gend1), Sigma = vcov(gend1), Terms = 6:15)
vif(gend1)
hoslem.test(genderdata$gender, fitted(gend1), g=10)

#Propabilities into arrays:
gendidea <- array(0, dim=c(2,2));
for (i in 1:2){
  gendidea[i, 1] <- 1/(1 + exp(-(coefficients(gend1)[2] *
   (i - 1) + coefficients(gend1)[1])));
  gendidea[i, 2] <- gendidea[i, 1] / gendidea[1, 1] - 1;
}

gendfin <- array(0, dim=c(6,2));
for (i in 1:6){
  gendfin[i, 1] <- 1/(1 + exp(-(coefficients(gend1)[3] *
   (i - 1) + coefficients(gend1)[1])));
  gendfin[i, 2] <- gendfin[i, 1] / gendfin[1, 1] - 1;
}

gendeng <- array(0, dim=c(6,2));
for (i in 1:6){
  gendeng[i, 1] <- 1/(1 + exp(-(coefficients(gend1)[4] *
   (i - 1) + coefficients(gend1)[1])));
  gendeng[i, 2] <- gendeng[i, 1] / gendeng[1, 1] - 1;
}

gended <- array(0, dim=c(6,2));
for (i in 1:6){
  gended[i, 1] <- 1/(1 + exp(-(coefficients(gend1)[5] *
   (i - 1) + coefficients(gend1)[1])));
  gended[i, 2] <- gended[i, 1] / gended[1, 1] - 1;
}

gendprof <- array(0, dim=c(11,3));
gendprof[1, 2] <- 1/(1 + exp(-(coefficients(gend1)[1])));
gendprof[1, 1] <- 0;
```

```
gendprof[2, 1] <- 1;
gendprof[3, 1] <- 10;
gendprof[4, 1] <- 20;
gendprof[5, 1] <- 30;
gendprof[6, 1] <- 40;
gendprof[7, 1] <- 50;
gendprof[8, 1] <- 60;
gendprof[9, 1] <- 70;
gendprof[10, 1] <- 80;
gendprof[11, 1] <- 90;
for (i in 2:11){
  gendprof[i, 2] <- 1/(1 + exp(-(coefficients(gend1)[4 +
   i] + coefficients(gend1)[1])));
  gendprof[i, 3] <- gendprof[i, 2] / gendprof[1, 2] - 1;
}

gendwork <- array(0, dim=c(9,2));
for (i in 1:9){
  gendwork[i, 1] <- 1/(1 + exp(-(coefficients(gend1)[16] *
   (i - 1) + coefficients(gend1)[1])));
  gendwork[i, 2] <- gendwork[i, 1] / gendwork[1, 1] - 1;
}
```

A.3 Appendix 3

The detailed distributions of all the professions and profession categories:

	None	62	2,0%
Industry and agriculture	Construction	415	13,2%
	Industry	42	1,3%
	Agriculture	70	2,2%
Artisanry	Carpentery	30	1,0%
	Clothing	92	2,9%
	Food making	144	4,6%
	Crafts	30	1,0%
	Metal work	24	0,8%
Technicians	Electrician	82	2,6%
	Plumber	15	0,5%
	Mechanic	157	5,0%
	Repairs	41	1,3%
Logistics	Car driver	76	2,4%
	Truck driver	41	1,3%
	Sailor	1	0,0%
	Machine driver	12	0,4%
	Logistic labor	40	1,3%
Health and care	Doctor	9	0,3%
	Nurse	33	1,0%
	Dentist	6	0,2%
	Elderly care	15	0,5%
	Child care	11	0,3%
	Pharmacy	11	0,3%
	Animal care	8	0,3%
Public officials	Teacher	90	2,9%
	Public official	18	0,6%
Security	Soldier	20	0,6%
	Police	18	0,6%
	Guard	33	1,0%
	Rescue	5	0,2%

Service work	Barber	65	2,1%
	Beauty	13	0,4%
	Waiter	86	2,7%
	Sales	115	3,7%
	Therapy	4	0,1%
	Customer serv.	33	1,0%
	Cleaning	161	5,1%
Specialists	Engineering	134	4,3%
	IT	132	4,2%
	Humanities	25	0,8%
	Architecture	13	0,4%
	Business	168	5,3%
	Design	37	1,2%
	Sciences	46	1,5%
	Law	35	1,1%
	Other	6	0,2%
Other	Student	144	4,6%
	Entrepreneur	82	2,6%
	Artist	43	1,4%
	Sports	31	1,0%
	Translation	17	0,5%
	Journalism	28	0,9%
	Media	11	0,3%
	Manager	65	2,1%
	Religious	4	0,1%

A.4 Appendix 4

The creation of the logistic regression model began by creating a model with all the predictors included. The following tables give details of the results of the different models on the way to the final model: predictors, their p-values/Wald test values, regression coefficients, GVIF values, and the entire model's AIC. Significant variables will be marked with a * and variables to be eliminated will be marked with an **E**.

The first model's results can be seen in the table A.1. There are already several statistically significant predictors ($p < 0,15$). GVIF values are not

causing problems. The first predictor to be eliminated is birth year, with a p-value of 0,88.

Predictor	β_i	**p-value**	**GVIF**
Intercept	-2,557	0,90338	
Database-ID	-1,59 e-4	0,01594*	1,286
Gender male	0,02278	0,38126	1,166
Birthyear	-1,573 e-3	0,88184**E**	1,251
Business	-0,07644	0,71132	1,808
Idea	-0,3397	0,16510	1,795
Finnish	0,2235	0,00408*	1,339
English	0,3283	1,89 e-6*	1,527
Nationality	Various	0,28	2,405
Education	0,1445	0,03378*	1,735
Profession	Various	0,0014*	2,097
Work exp.	0,2702	3,71 e-10*	1,142
AIC	1411,6		

Table A.1: First model with all predictors included

The second model was created without the birth year variable, see table A.2. The next variable for elimination was one's willingness to establish a business, with a p-value of 0,71.

Predictor	β_i	**p-value**	**GVIF**
Intercept	-5,685	2,02 e-7*	
Database-ID	-1,598 e-4	0,01581*	1,287
Gender male	0,2261	0,38427	1,164
Business	0,07744	0,70756**E**	1,806
Idea	-0,3382	0,16652	1,792
Finnish	0,2216	0,00391*	1,305
English	0,3269	1,65 e-6*	1,497
Nationality	Various	0,27	2,225
Education	0,1458	0,03089*	1,707
Profession	Various	0,0014*	2,050
Work exp.	0,2708	2,91 e-10*	1,133
AIC	1409,6		

Table A.2: Second model's results

The third variable to be eliminated was gender, with the p-value of 0,39, see table A.3 for details.

Predictor	β_i	**p-value**	**GVIF**
Intercept	-5,705	1,77e-7*	
Database-ID	-1,561e-4	0,01711*	1,257
Gender male	0,2245	0,38756**E**	1,163
Idea	-0,3950	0,03906*	1,098
Finnish	0,225	0,00376*	1,305
English	0,3269	1,64 e-6*	1,496
Nationality	Various	0,26	2,196
Education	0,1450	0,03189*	1,704
Profession	Various	0,0014*	2,0046
Work exp.	0,2699	3,17 e-10*	1,130
AIC	1407,7		

Table A.3: Third model's results

The fourth variable to be eliminated was the first multicategorical variable: nationality, see table A.4.

Predictor	β_i	**p-value**	**GVIF**
Intercept	-5,522	2,49 e-7*	
Database-ID	-1,595e-4	0,01451*	1,254
Idea	-0,3908	0,04111*	1,097
Finnish	0,2165	0,00463*	1,294
English	0,3333	8,62 e-7*	1,480
Nationality	Various	0,26**E**	2,030
Education	0,1428	0,03428*	1,701
Profession	Various	0,001*	1,953
Work exp.	0,2730	1,83 e-10*	1,125
AIC	1406,5		

Table A.4: Fourth model's results

The model without nationality turned out to be the final model, presented in section 5.2.

A.5 Appendix 5

The creation of the logistic regression model for the predictors began by creating a model with all the predictors left out from the main model. The following tables offer insight to each model created during the building process: predictors, their p-values/Wald test values, regression coefficients, GVIF values, and the entire model's AIC. Significant variables will be marked with a * and variables to be eliminated will be marked with an **E**.

The first model's details can be seen in the table A.5. The first variable to be eliminated will be the willingness to establish a business, with its p-value of 0,41. So far only gender seems to have a significant indirect connection.

Predictor	β_i	**p-value**	**GVIF**
Intercept	-22,39	0,2460	
ID	6,722 e-5	0,2533	1,133
Gender male	0,5043	0,0381*	1,085
Birthyear	9,572 e-3	0,3232	1,105
Business	-0,1256	0,4111E	1,060
Nationality	Various	0,067	1,314
AIC	1517,2		

Table A.5: Results of the first predictor model

The second variable to be eliminated is birthyear, with a p-value of 0,32, see table A.6 for details.

Predictor	β_i	**p-value**	**GVIF**
Intercept	-22,70	0,2392	
ID	7,620 e-5	0,1873	1,047
Gender male	0,4968	0,0410*	1,084
Birthyear	9,695 e-3	0,3167E	1,104
Nationality	Various	0,067	1,281
AIC	1515,9		

Table A.6: Results of the second predictor model

The third eliminated variable will be the control variable database-ID, with a p-value of 0,17, see table A.7.

Predictor	β_i	**p-value**	**GVIF**
Intercept	-3,396	¡2 e-16*	
ID	7,861 e-5	0,1733E	1,094
Gender male	0,5096	0,0358*	1,081
Nationality	Various	0,081	1,163
AIC	1515		

Table A.7: Results of the third predictor model

The fourth variable to be eliminated is nationality, with a p-value of 0,06, see table A.8 for reference.

Predictor	β_i	**p-value**	**GVIF**
Intercept	-3,234	<2 e-16*	
Gender male	0,4899	0,0434*	1,075
Nationality	Various	0,064E	1,075
AIC	1514,8		

Table A.8: Results of the fourth predictor model

The final model to be analyzed turned out having only one predictor: gender. This model is showcased in the section 5.3.